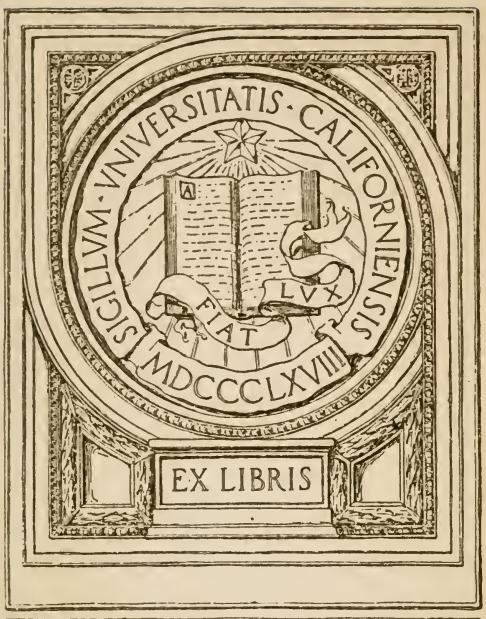


BIOLOGY
LIBRARY
EDUC.
PSYCH.
LIBRARY

JUDGING HUMAN CHARACTER

BY HARRY L. HOLLINGWORTH

JUDGING HUMAN CHARACTER

THE PSYCHOLOGY OF FUNCTIONAL NEUROSES

ADVERTISING AND SELLING
Principles of Appeal and Response

VOCATIONAL PSYCHOLOGY

APPLIED PSYCHOLOGY
(With A. T. Poffenberger, Jr.)

D. APPLETON AND COMPANY
Publishers New York

T216 C

JUDGING HUMAN CHARACTER

BY

H. L. HOLLINGWORTH

PROFESSOR OF PSYCHOLOGY, BARNARD COLLEGE, COLUMBIA UNIVERSITY; AUTHOR OF "ADVERTISING AND SELLING: PRINCIPLES OF APPEAL AND RESPONSE," "VOCATIONAL PSYCHOLOGY," "PSYCHOLOGY OF FUNCTIONAL NEUROSES," ETC.



D. APPLETON AND COMPANY
NEW YORK :: 1922 :: LONDON

BF 818
H6
BIOLOGY LIBRARY
EDUC.
PSYCH.
LIBRARY

COPYRIGHT, 1922, BY
D. APPLETON AND COMPANY

NO. MIN.
PRICE

PRINTED IN THE UNITED STATES OF AMERICA

TO

A GROUP OF BOYHOOD COMRADES

E. LeGRANDE CHERRY, C. LeROY CHERRY,
J. REID GREEN, GLENN N. VENRICK,
EDW. L. WITTE

483122

PREFACE

The following chapters present, in organized arrangement, part of the material included in a series of lectures on vocational and industrial psychology. Only those topics are considered which bear more or less directly on the appraisal of human character, for definite and usually for practical purposes. The book does not undertake to outline plans for vocational guidance, nor to prescribe specific procedures in personnel work. It offers, instead, a general and nontechnical survey of the principle methods relied on in judging human traits.

Traditional methods are criticized, and effort is made to suggest improvement in the technic of applying these methods or in interpreting the results derived from their use. The present status of the methods of mental measurement in various fields is surveyed, in the endeavor to communicate a general understanding of the principles underlying these methods. No attempt is made to present nor to advocate special tests or scales in any of the fields considered, but references are given in the bibliography which should

serve to lead the reader to a more detailed acquaintance with the special methods in the fields of his own particular interests.

The book is not, therefore, intended primarily for the professional psychologist, but for the general reader, for the student, the teacher, the business man, foreman, executive, or personnel specialist, to whom the technical journals are unfamiliar, inaccessible or unintelligible.

Readers familiar with the writer's previous publications in this general field, especially with *Vocational Psychology* (Appleton, 1916) and *Applied Psychology* (with A. T. Poffenberger, Appleton, 1917) will observe that some of the tables of data previously published are here made use of again. Such results, however, have usually been extended, modified or submitted to more detailed interpretation and application. The last six chapters also deal with topics considered in the earlier publications. These are topics which have undergone considerable development during the years that have elapsed since the earlier publication, and the present chapters serve to indicate the general nature and direction of that progress. In some ways the present volume may be considered supplementary to the earlier book,

Vocational Psychology, in so far as it surveys the contemporary status of some of the problems or methods indicated in that book.

Systematic attempts to measure and appraise the qualities of men and women play a conspicuous rôle in the program of modern experimental psychology. Even when methods and results lack completeness and finality, it is well to bring together the various lines of development in the form of a survey of present accomplishment. To give such a survey was the purpose of the course of lectures which served as the basis for the chapters of this book. Two appendices have been added. In one is given a selected bibliography arranged under the chapter headings of the book. In the other is presented a series of laboratory exercises which have been found useful for demonstration purposes or for general class use in connection with courses in Mental Measurement, Vocational Psychology, Employment Practice, and similar topics.

To Prof. W. V. Bingham, Prof. A. I. Gates, and Dr. Georgina S. Gates, the author is indebted for valuable suggestions in connection with the preparation of the manuscript.

H. L. Hollingworth

Barnard College, Columbia University

CONTENTS

CHAPTER		PAGE
I.	DIAGNOSING HUMAN CHARACTER . . .	1
	Judgment and Perception—The Data of Judgment—Traditional Methods.	
II.	LETTERS OF APPLICATION	9
	An Experimental Test—Experimental Results—Interpretations—Improved Technic.	
III.	HUMAN CHARACTER IN PHOTOGRAPHS . .	33
	Agreement of Judges—Unreliability of Individual Judgment — The Consensus of Opinion—Additional Evidence.	
IV.	JUDGING ONE'S OWN CHARACTERISTICS .	45
	Accuracy of Self-Estimates—Constant Errors in Self-Estimation—Qualification of Judges—Improved Technic.	
V.	THE PERSONAL INTERVIEW	60
	Experimental Tests of Interviews—Similar Investigations—Improvements in Method.	
VI.	RECOMMENDATIONS AND TESTIMONIALS . .	74
	Testimonial Disagreement—The Hierarchy of Consistency—Objective and Subjective Traits.	

CHAPTER PAGE

VII. Common Sources of Error and Their Correction 91

Special Tendencies of Judgment—Testimonial Validity — The Elimination of Variability—Rating Scales—A Model Inquiry Form—Suggested Improvements.

VIII. Measuring Mental Competence . . 112

Tests and Measurements—Principles of Construction—Principles of Expression—Illustrative Test Procedures—Analogies Test—Primary Meaning of Test Records—Choice of Tests—Individual and Group Methods.

IX. General Competence and Special Aptitude 139

Specialized Traits—An Experimental Demonstration — Temperamental Qualities — Illustrative Cases.

X. The Method of the Sample 151

The Procedure of Sampling—Trade Test Methods—Educational Measurement.

XI. Character Analysis through Correlation 161

Empirical Correlation Procedure—A Typical Investigation—Differential Tests.

CHAPTER PAGE

XII. The Diagnosis of Temperament . . . 179

Moral and Social Traits—The Psychoneurotic Inventory—The Significance of Handwriting—Tests of Aggressiveness—The Association Methods—Measurements of Suggestibility—Detecting Moods and Emotions—General Conclusions.

XIII. Psychographic Methods 201

Psychographic Technic—Typical Psychographic Profiles—Uses of the Profile Method.

APPENDICES

PAGE

A. Bibliography 215

B. Experiments and Laboratory Exercises . 225

EXERCISES

I. Letters of Application 226
II. Human Character in Photographs . . 228
III. Judging One's Own Characteristics . . 231
IV. The Personal Interview 234
V. Recommendations and Testimonials . . 237
VI. Rating Scales 239
VII. General Principles of Tests 242
VIII. Measuring General Competence . . . 244
IX. Measurement of Special Aptitudes . . 246
X. Measures of Knowledge and Skill . . . 248
XI. Educational Measurement 250
XII. The Correlation Methods 252
XIII. Temperamental Characteristics . . . 256
XIV. Psychographic Methods 258
XV. Examination of an Individual 260
XVI. Installation of Employment Supervisor . 262

Index 265

JUDGING HUMAN CHARACTER

CHAPTER I

DIAGNOSING HUMAN CHARACTER

The diagnosis of human character is an enterprise in which all men, women and children engage. From infancy to senility we are impelled to attempt it, and to venture our own hopes on the accuracy of our judgment. Personal adjustments, family life, social companionship, business relations, politics and diplomacy alike depend on more or less accurate knowledge of the qualities of men. Even a dog, to be happy, must learn to discriminate between the signs of friendliness and those of an evil intention.

Judgment and Perception

In most situations, however, it is the perception of character, rather than the judgment of it, on which we rely. The distinction between these two methods, though not to be sharply drawn, is

nevertheless essential. Through repeated and prolonged experience with another, in varied circumstances and under varied provocations, we come to *know* his character. For by character we mean essentially the characteristic modes of behavior, the characteristic attitudes, reactions and capacities. Prolonged acquaintance gives as direct a perception of character, in this sense, as it does of the qualities of foods and climates or the properties of physical objects.

Judgment of character, as distinguished from its perception, is a more indirect process, in which, from a momentary observation or a cross section of the life of another, we attempt to estimate its general quality and tenor. Judgment, in this sense, is diagnosis. It relies on symptoms, signs, clues and incomplete evidence. It is an inference from the quality of a detail to the quality of the larger whole.

Whether we judge or whether we perceive character will thus depend in part on the degree of acquaintance—on the amount of information at our disposal. The total stranger and the individual only occasionally seen, we must judge. The character of our more intimate associates we know directly, from having more fully experienced

it. For the character of a man is not some hidden substance or possession, nor is it a mysterious spiritual essence. A man's character is his actual behavior, when all of his conduct is considered. Lovableness is just the degree to which people are fond of us; kindliness and benevolence are present to just the degree that people are actually gratified and comforted by our conduct. Just as the value of a commodity depends on what people will actually pay for it, so human traits are constituted by the degree to which they are actually manifest. A single bid at an auction may not tell the whole story of an article's worth, nor an isolated symptom disclose the whole of a patient's disease; neither may the single act of a man reveal his whole character. But the final bid and the differential symptom have high significance. So also may some of a man's acts be more significant than others.

The earlier, in our acquaintance with another, we seek to formulate and express his temperament, interest or ability, the more we must rely on judgment and the less on full perception. Judgment, in the sense in which we use the term, is an indirect approach to knowledge through the interpretation of signs, cross sections, or partial

details. In many human relations this type of character estimate must be relied on, since full acquaintance is lacking, or since action must be taken before full acquaintance can be attained. Especially is this the case in business relations, in which people must be selected from a larger number, employed for particular work, promoted to positions of responsibility, consulted concerning matters of moment. In general, any executive dealing with large numbers of individuals can know these individuals only indirectly and through their occasional acts. Operators must be hired, apprentices chosen, clerical workers selected, instructors provided, to handle increased volume of work or to take the place of those who are dismissed, promoted or retired. In all such cases the executive or foreman must exercise his judgment, and must exercise it upon such data as are available. It is our purpose to examine somewhat closely and experimentally, this judgment and these data.

The Data of Judgment

It should no longer be necessary to recapitulate nor to expose the many quaint and bizarre results of the very earliest attempts to systematize the

signs of character. The motivation and the inadequacy of these curious efforts have been exhibited in other places.[1] The magic of primitive men sought to *produce* traits and attitudes in others, rather than to *discover* them. Wishes may indeed go far in determining the structure of human minds, but they are neither prompt, certain, nor sure enough to satisfy the demands of an employment manager, a teacher, a foreman, a parent or an executive. When fortune-making became fortune-telling, the signs and details of destiny were sought quite outside the individual's own conduct, in stars, omens and birthdays. In time these indiscriminate clairvoyances also gave way, and the field of significant signs was limited more generally to the structural characteristics of the individual. His shape, posture, gait and pigmentation, his contours, wrinkles, protuberances and asymmetries, all these were hopefully believed to present a system of signs and clues to his feelings, thought and conduct.

Only generations of disappointment, chagrin and absurdity have sufficed to lead men to seek for the signs of conduct in conduct itself, and to

[1] See *Vocational Psychology* (by H. L. Hollingworth, D. Appleton and Co., New York, 1916), especially Chapters I and II.

place more reliance on an actual sample of behavior than on a presumed astrological or anatomical sign of it. It is at the point of this shift in interest that our own discussion is to begin. We are to consider those judgments of character which are based on conduct itself, and shall first consider a group of *traditional* methods of judging character and aptitude. This will be followed by the account of more modern endeavors to place character judgment on a basis of verifiable fact and principle.

Traditional Methods

The father of a lively young grammar school boy once announced that he was preparing to have his son taught pharmacy. When asked the reason for this definite and early vocational direction of the youth, he replied that he had taken great pains to observe the behavior of his son ever since his infancy, and that he had decided to make a druggist of him because "as a child he was so fond of pouring things from one bottle into another."

During the recent war a certain army hospital required the services of a pharmacist, and one was duly requisitioned from headquarters. In

time the new pharmacist reported for duty—an awkward, strapping fellow, with a burly frame and dirty hands. "Are you a pharmacist?" he was asked by the officer in charge. "I sure am, sir," he replied. "What drug store did you work in?"

"Oh, I never worked in no drug store." "You never worked in a drug store and yet you say you are a pharmacist!" "Oh, yes, sir, I'm a right good farmacist. I worked on a farm all my life."

A contemporary vocational counselor is said to receive his clients in an office with no hat rack and no extra chair. The client enters and the counselor abruptly orders him, "Hang up your hat! Sit down!" The amazed young chap, seeing no place to hang his hat, and finding no chair to sit on, does one of two things: (1) He may resent the insult, register anger, and perhaps make a justly impudent remark. (2) He may be so astounded by the counselor's unexpected impertinence that he stands awkwardly fumbling his hat, grinning or blushing, with a demeanor that is at least outwardly meek and humble. In the former case he is advised to become a salesman; in the latter he is advised to take up the study of pharmacy, or some similar vocation.

These authentic illustrations are not designed

to bring ignominy upon the drug business. Instead they serve to introduce some of the traditional methods of judging character, aptitude, and interest, based as they are on: (1) the observations of a tender parent; (2) the candidate's statement of his own qualifications; and (3) the impressionistic theory of a prejudiced and ignorant interviewer. We may add to these methods three further traditional procedures: (4) The letter of application; (5) the photograph; and (6) the recommendation. Perhaps the last three methods are to-day in better popular repute than are the first three. But whatever their repute, their adequacy deserves investigation, and their importance merits any improvement that can be made upon them. In the sections immediately following, accounts are given of various investigations directed toward the evaluation and improvement of several of these traditional methods.

CHAPTER II

LETTERS OF APPLICATION

Under modern conditions the personnel of a business or institution comprises so many different individuals, and activities are carried on in such large centers or in such remote districts, that employees and assistants must often be chosen without the personal acquaintance of the superintendent or executive. At least the initial applications for a given place are often so numerous and submitted from so great a distance that even a personal interview with all candidates is not feasible. Under such circumstances it is common practice to require each candidate to submit a letter of application, in which his or her qualifications are set forth and certain personal data recorded. Sometimes a standardized or conventionalized application form is used. More frequently the form and in part the content of the letter is left to the wisdom and discernment of the applicant. Through the inspection of these

letters of application many candidates are at once rejected and not further considered. Other letters may lead to more favorable action—the applicants may be requested to call for a personal interview, or may even be judged as suitable on the basis of the letter alone.

An Experimental Test

Our present task is that of examining somewhat more closely than is usually done the judgments of human character that are based on such letters of application. In order to secure representative material for the study, a *bona-fide* advertisement for a bookkeeper and office assistant was inserted in the "help wanted" columns of the Sunday editions of two New York City newspapers. Over one hundred letters of application were received. Each gave, in the applicant's own handwriting, and on stationery individually chosen, the main facts of the applicant's business career, education, experience, and previous employment, and set forth with such clearness as the applicant could command the particular qualifications for the position in question. From this large number of letters twenty-five were chosen at random to

serve as material for an experimental investigation.

These letters, each bearing a key number to aid in its identification, were presented to fifty different judges. One group of judges consisted of business men and women who were constantly being called on, in their actual affairs, to make such judgments. Another group consisted of professional men and women, including a number of psychologists; another group consisted of miscellaneous individuals, students, clerical workers, etc. The judges were instructed to place themselves in the position of a prospective employer and to arrange the twenty-five letters in order of merit, as replies to the given advertisement, with respect, first, to the intelligence; second, the reliability; third, the tact; and finally, the neatness, indicated by the letters. Comparison of these different arrangements will at least reveal the amount of agreement shown by the estimates of different judges. If the disagreement is great, it will be evident that the estimates of a single judge are unreliable.

As a further test of the stability of such judgments, ten of the judges were again approached

after a month had elapsed and requested to arrange the letters again, for the same traits and from the same point of view as on the previous occasion. This makes it possible to determine how consistent are the estimates passed by a given judge, and how far his estimates vary with lapse of time, change of circumstance, or a shift of mood and disposition. If a given judge passes quite different verdicts on two different occasions, on precisely the same material, it is clear that any one of his judgments is likely to be unreliable. The agreement of different judges and the consistency of given judges thus afford criteria for the evaluation of character estimates based on letters of application.

Experimental Results

The estimates, some of them repeated, by fifty judges, on twenty-five letters, of four different traits, give a mass of figures entirely too unwieldy to repeat here.[1] We shall choose for presentation typical results by taking from the complete tables the estimates of the first ten judges for ten of the

[1] The complete data have been recorded by Lillian C. Walton, who conducted this investigation, in her Master's essay, entitled "A Study of Judgments of Letters of Application," on file in the library of Columbia University.

letters, using always the same judges and the same letters. The position given a single letter may be anywhere from 1 to 25, position 1 being the best for the trait in question, and 25 being the poorest.

ESTIMATES OF TEN LETTERS BY TEN JUDGES

Intelligence

Judge	A	B	C	D	E	F	G	H	I	J
I	6	24	13	20	5	3	14	19	12	11
II	13	15	6	2	5	16	14	17	12	18
III	2	17	5	22	9	6	13	21	23	14
IV	11	22	18	13	19	8	20	25	9	16
V	9	25	19	20	3	5	18	13	16	14
VI	17	14	25	12	22	3	5	21	20	19
VII	3	5	9	7	13	1	10	24	15	11
VIII	4	14	12	17	6	10	13	16	21	22
IX	11	4	7	18	16	3	5	17	19	23
X	3	20	9	19	5	8	22	17	18	16

Tact

Judge	A	B	C	D	E	F	G	H	I	J
I	10	24	21	14	7	4	23	11	13	19
II	17	6	4	5	2	16	24	25	7	19
III	2	23	3	22	4	11	17	16	21	13
IV	23	18	25	16	3	7	10	13	15	22
V	10	11	21	19	13	4	12	2	22	5
VI	13	20	10	16	11	5	2	18	17	23
VII	5	9	3	10	4	1	6	22	14	23
VIII	4	5	12	13	6	10	14	16	21	22
IX	10	4	7	18	16	3	5	22	25	21
X	16	1	2	14	20	17	21	7	15	13

Reliability

Judge	A	B	C	D	E	F	G	H	I	J
I	6	16	7	19	14	3	5	18	11	21
II	17	6	5	9	7	19	24	25	3	21
III	9	18	4	21	16	12	17	11	7	23
IV	17	18	24	23	1	15	16	25	4	10
V	19	20	2	14	12	17	21	1	3	4
VI	12	17	9	19	11	2	4	21	15	23
VII	5	13	25	6	8	1	2	21	14	12
VIII	4	14	12	17	6	10	13	16	21	22
IX	11	4	7	18	16	3	5	17	19	21
X	5	20	14	18	19	2	15	11	12	7

Neatness

Judge	A	B	C	D	E	F	G	H	I	J
I	7	22	14	21	11	15	4	16	9	20
II	13	21	5	22	12	14	24	18	3	17
III	1	20	3	24	5	7	16	17	8	14
IV	11	25	18	14	8	17	15	16	6	23
V	13	25	6	22	2	9	10	7	20	21
VI	8	13	5	23	7	10	3	19	15	24
VII	11	13	7	21	6	2	9	8	3	23
VIII	4	13	12	16	6	10	14	17	21	22
IX	10	4	7	18	16	3	5	17	19	23
X	9	22	14	18	5	8	20	15	7	17

Interpretations

These tables, although they do not show the whole range of variation in the judgments, are sufficient to suggest that this variability is almost as great as it could possibly be. In the case of Neatness, for example, the letter marked "A" for

purposes of identification was given the highest place (1), the lowest place (25), and also occupied positions all along the scale, from best to poorest. Letter "B" was placed as high as position 4, as low as position 25, and was assigned positions all along the scale by various judges. Letter "C" ranged from 2d to 25th place, and most of the other letters produced similarly scattering verdicts. This is true for Neatness, in spite of the fact that it is in judging this trait that the different judges agreed most closely.

If the letters had been arranged in purely chance orders they would have had much the same average position and each letter would tend to occupy each position in the scale equally often with all other letters and all other positions. And the chance deviation of the various positions from this average would be a little over six steps. Under the circumstances of the experiment the arrangements made by the various judges approach very closely to chance series. The average deviations of the letters from their own average positions are for Neatness 4.83 steps, for Intelligence 5.52 steps, for Reliability 5.82 steps, and for Tact 6.23 steps. Only Neatness and Intelligence show agreement that is at all certainly

closer than that which chance shufflings of the letters would have produced. Even here the tables show that the letter that one judge would have carefully filed for future reference in a personal interview with the applicant, another judge would without further consideration have thrown into the waste basket. Certainly a method that results in such disagreement cannot be justified merely on the grounds of its simplicity, cheapness, and antiquity. Its use results in unfairness both to the judge and to the applicant.

The facts are still more striking when we consider the consistency with which the same judge reacts on different occasions, although judging precisely the same letters, after having in the main forgotten his previous verdicts. Here we may make use of a method of measuring resemblance through statistical formulae. Mathematical treatment of the figures can be made to yield coefficients of similarity between two such arrangements of materials. Purely chance agreement will on this basis yield a coefficient of .00. Complete agreement will give a coefficient of + 1.00, and completely reversed arrangements will give — 1.00. The various degrees of similarity or difference will give coefficients ranging anywhere

from +1.00 through .00 to −1.00. Thus a coefficient of +.50 would represent resemblance of two series part way between perfect similarity and purely chance agreement. These measures may be given a more concrete meaning by reference to the resemblances of people. Children of the same family resemble each other in physical appearance more than do children chosen at random. And twins resemble each other still more closely than do brothers or sisters who are not twins. Studies have been made in which children have been measured in various physical characteristics, and the degree of resemblance stated in just such coefficients of similarity as those we have described. When this is done it is found that children of the same family give coefficients of resemblance of about +.40. Twins, however, give coefficients of about +.80. We may then say, by way of concrete illustration, that judgments that agree with each other by a coefficient of .40 are as similar as children of the same family, while judgments that give coefficients of similarity as great as .80 are as alike as twins.[2] The follow-

[2] Readers interested in the detailed statistical methods used in such computations will find them discussed in most of the modern books on statistical methods. It is well to have the concept fairly clearly in mind since the convenience of such mathematical expressions will lead us to make considerable use of them.

ing table gives the coefficients of correlation in the case of the ten judges who arranged the letters on two different occasions. These correlations measure the consistency of the two trials.

COEFFICIENTS OF CORRELATION

Judge	Intelligence	Tact	Reliability	Neatness	Average
2	.59	.40	.50	.67	.54
3	.72	.72	.73	.72	.72
18	.08	.40	.27	.38	.28
5	.72	.44	.65	.88	.67
20	.60	.63	.20	.44	.47
39	.31	.18	.23	—.14	.21
7	.44	.52	.46	.92	.60
6	.62	.31	.45	.51	.47
15	.65	.71	.73	.91	.75
1	.63	.42	.52	.71	.57
Medians	.61	.43	.48	.69	.55

The results show clearly that two such arrangements on different occasions are far from being as similar as twins. Neatness and Intelligence again give the most consistent results, the median coefficients being + .69 and + .61. Tact and Reliability give median coefficients of + .43 and + .48, just about as much similarity as exists between children of the same family in general. If these coefficients had shown such resemblance as one finds between twins, the judgments might

very well have been considered stable and characteristic of the various judges. But such coefficients as we actually secured indicate that on the whole a given judge does not resemble himself any more than he resembles other judges. And we have already had occasion to see what diverse results can come from shifting from the verdict of one judge to that of another.

The table further shows that some judges are more consistent in their ratings than are others. Thus judges 3, 5, and 15 have fairly high coefficients for all the traits, their different arrangements being almost as similar as twins. Judges 18, and 39 on the other hand have very low coefficients of consistency, and indeed judge 39 in the case of Neatness tends toward verdicts that are the reverse of those of his earlier arrangement, thus giving a negative coefficient (—.14).

In much the same way, if we take as our standard the average arrangement of all the judges, their combined opinion, and compare each judge's arrangements with this standard, we find that some judges are more in accord with the group average than are others. Some judges agree closely with this standard in all of the traits. Some differ from it markedly in all traits.

The majority, however, agree with the standard to various degrees in the case of the different traits. There is so much disagreement on the whole that it is quite impossible to pick out any one of the judges as a "general expert," that is, one who in judging all of the traits gives arrangements that approximate very closely to the consensus of opinion of all the judges. Nor do the business men and women, who have had experience in passing on such letters agree any more closely with this consensus of opinion than do the psychologists or the miscellaneous individuals.

A study of twenty-five letters of application has been reported by Poffenberger and Vartanian,[8] which confirms many of the results we have just considered. The letters were written by seniors in a training school for religious workers, as applications for a position in the kind of work for which they were in preparation. No application form was used, the letters being in the handwriting of the candidates, and giving what each considered to be the relevant or significant details of his characteristics, qualifications, experience and training. "These letters just as received were

[8] Poffenberger and Vartanian. "The Letter of Application in Vocational Selection," *Journal of Applied Psychology*, March, 1922.

given in turn to twelve members of the staff of the Union Theological Seminary with the request that they arrange them according to the degree to which they indicated general fitness for the position."

Criteria for checking up the validity of the judgments of the letters were secured in the following way: "Five teachers from the staff of the training school furnished three separate arrangements of the twenty-five individuals according to the degree to which they possessed the three traits, general ability, intelligence and tact. In addition each member of the group of applicants arranged his twenty-four associates and himself in an order for each of these three traits. The conditions for such estimates were especially good, as the school is small and every teacher has very close association with his students; and the students themselves are more closely associated than are students in the ordinary college."

In this case, although it is not possible to compare the judgments of applications with strictly objective data, they can be compared with judgments based upon actual acquaintance, recorded by teachers and by fellow students. When in each case the group judgment or consensus of

opinion is taken as the measure, the following correlations result.

CORRELATION OF GROUP JUDGMENT

Group Judgment of Letters Correlated with	General Ability	Intelligence	Tact
a—Estimates of Fellow Students....	.46	.44	.18
b—Estimates of Teachers.....	.56	.58	.20
c—Students and Teachers Combined.	.50	.44	.22

In the case of Tact the correlations are so low as to be very unreliable, and it will be remembered that Walton also found both consistency and agreement to be least in the case of that trait. In the case of General Ability and Intelligence, however, the correlations average .50, indicating a very real tendency to agreement between group judgment of letters, by strangers, and group estimates of personal traits by acquaintances.

Considering the individual judges of the letters, when the estimates of fellow students are taken as the standard, the average of the individual correlations is .37 in the case of General Ability. The individual correlations range from .18 to .52. When the estimates of teachers are taken as the standard, the individual correlations of the judges range from .24 to .57, averaging .43. That is to

say, the group judgment of the letters of applications accords better with the estimates of acquaintances than do the individual judgments, on the average. "Although one might find a judge who would do better than the group judgment, the group judgment would be safer unless one had some means of knowing the good judges beforehand."

Improved Technic

Such results do, however, suggest a useful procedure in selecting and in checking up the individuals who are assigned the work of passing judgment on letters of application, when such applications must be considered. If the consensus of opinion be derived not from the judgments of random individuals, but from a number of members of the firm, alike conversant with the qualities desired, with the definition of trait terms, and with the general policy of the firm, the judicial capacity of each may be determined. This will be shown by the individual's agreement with the consensus of opinion of all the responsible individuals. Deviation from this standard will indicate at least departure from the general aim or policy. Close agreement with the consensus

will characterize the executive whose judgments best represent the combined opinions of the firm. In the case we have just considered, judges 3, 5, and 15 are the ones best qualified to represent the group opinion, if we assume that all four of the traits here considered are essential. On the other hand it would be folly to assign the task of rating such applications to judges 18 and 39, whose agreements with the consensus of opinion are so low as to suggest almost purely random decisions, unless it could in some definite way be demonstrated that the policy of the firm or the consensus of opinion were wrong. This could be learned by comparing the early ratings with the subsequent success of such candidates as were employed. Such investigation requires time and patience, as well as considerable experimental and statistical expertness, but such procedure represents the only promising method for giving value and accuracy to judgments based on letters of application, and for choosing employment executives whose opinions, in this connection, will have substantial merit.

A further step with genuine utility consists in taking care that the application will contain data

of the highest possible relevance. Shall the applicant be required to state: (*a*) his mother's maiden name; (*b*) whether his father is living; (*c*) his reason for leaving school; (*d*) what studies he liked best; (*e*) what church he belongs to; (*f*) whether he plays a musical instrument; (*g*) how many brothers and sisters he has; (*h*) whether he is a vegetarian; (*i*) his pulse rate? All of these are questions that actually occur on application forms. Obviously such an application or letter should contain as much relevant material as possible and little distracting and irrelevant material. But the preparation of instructions for such applications presupposes that some study has been made of the relation between biographical data and fitness for the job in question. Furthermore, not all relevant biographical facts are equally important for many jobs. In one connection age may handicap while education qualifies; in another connection education may count for little but physiological maturity may count for much.

Judgments based on letters of application may therefore be given added validity if the letters are required to give relevant data and if there is some more or less systematic method of evaluating the relative importance and the total significance of

the data reported by the candidate. No precise rules can be laid down for such an enterprise in general, since the value of every item will vary with the circumstances, the job, the firm, and with other items. But for any particular situation investigation will usually disclose valuable principles or tendencies, and often exact quantitative measurements of the value of special items may be secured. The following examples will suggest ways in which added value may accrue to the traditional method of selection based on letters of application.

The National Association of Life Underwriters recently secured, from 548 successful insurance salesmen, information on some 70 questions, partly of a biographical character. An analysis of these returns has been made by Dr. F. L. Wells.[4] The men are grouped according to earning capacity, into four groups which we may roughly characterize as Excellent, Good, Fair and Poor. On the basis of annual earnings they represent $25,000, $10,000, $5,000 and $2,500 groups. From the numerous items given in the analysis of Wells, the following are chosen merely to illustrate the

[4] F. L. Wells. "Analysis of a Successful Agent," *Life Assoc. News*, XI, 3.

difference between relevant, ambiguous and irrelevant facts.

EARNING CAPACITY AND BIOGRAPHICAL DATA

	Excellent	Good	Fair	Poor
Number of men in the group...	8	81	180	279
Average age.....	46	45	44	41
Percentage of time given to work	88	77	74	78
Number married (per cent)...	90+	90+	90+	90+
Per cent having children.....	100	82	74	67
Per cent working as a boy.....	50	80	84	85
Per cent starting on commission	88	86	81	37
Per cent having good health...	100	97	94	95
Attracted to work by opportunities	50	60	62	60
Entering work, from necessity..	25	6	12	14
High school or seminary education	100	65	68	67
Attended college or technical school	63	27	36	73
Per cent having recreations....	100	87	82	74
Having mortgage on own house	13	30	30	39
Feeling lack of perseverance...	0	1	9	17
Reporting lack of system.....	0	11	2	6

All of these items represent data that could be secured from a properly prepared application blank or letter. But it is clear that they are not of equal importance, from the point of view of success in the occupation concerned. Thus Time Given to Work, Having Good Health, Motive for

Entering Work, Education, Felt Lack of System, and Marital Status, do not vary with earning power or success in this vocation, especially if the very small group of eight "excellent" men be omitted because of the relative unreliability of averages from so small a number of cases. Age, Working as a Boy, Mortgage on House, tend only slightly to be related to success.

On the other hand Having Children, Starting Work on Commission, Having Outside Recreations, Not Feeling Lack of Perseverance are definitely related to success, throughout the table of results. Now if, on general grounds, health, education, marriage, freedom from mortgages, and system are desirable traits in an applicant, it would seem that these traits should at least be given less weight than that assigned to those traits that are definitely related to degree of success. To assign precise weights to various items involves somewhat complicated technic, such as that to be described in a later chapter on the methods of correlation. But an improvement over the random method commonly used would be to assign, say, one point credit to desirable traits not found to be related to success, and two points to traits or items found to be so related. By some

such system of point credits an approach to objective ratings of application blanks may be made which would quite certainly possess a validity not characteristic of judgments based solely on subjective impression.

Andrews[5] has reported, from the Bureau of Personnel Research of the Carnegie Institute of Technology, a case in practical business in which such a system of grading applications was adopted, after a preliminary study of the relation between various biographical facts and success in the case of individuals already in the employ of the firm. The following quotation illustrates the use of the method.

A man states in his application that he is thirty-five years old. Looking up the range of values for age we find that he gets three points on that item. We note that he is single and we take away a point, leaving two. He has had eight years schooling and we add a point for this, bringing his total thus far back to three. He has been a file clerk, ledger clerk, and head clerk, all of which come under social occupations; we subtract a point from the total. He has, however, remained for years with the same company, which shows that he is a stable worker and not a "floater." This adds a point and his total once more stands at three. After assigning

[5] L. G. Andrews. "A Grading System for Picking Men," *Sales Management*, January, 1922.

the proper scores to the remainder of the items on his application blank, we finally secure, let us say, a total score of 8. Reference to our table shows that men scoring 8 have practically always made good, so the man is hired.

While such improvements of the traditional method cannot be expected to produce perfect correspondence between selection and success, the evidence is that the careful scrutiny, system and objective point of view which the method involves constitute a genuine advance in personnel selection. Thus Andrews concludes that "the items in an application blank can be graded so that they will pick the right man 60 per cent of the time."

Similar methods of analysis were used by Thorndike [6] in his study of the biographical data of applicants for the schools of military aeronautics. Biographical items were compared with the classification of men as successes or failures in the Ground Schools, the Flying Schools, and in actual service, in the hope of discovering significant antecedents which might aid in the selection of men for training. In this way the most important items were found to be amount of edu-

[6] *The Personnel System of the U. S. Army*, vol. i., pp. 604-633.

cation, rate of school progress, class standing, and interest in and ability at studies of the physical sciences, when success in the Ground Schools was taken as the criterion. Age, salary at last position, social status (occupation of father) showed no correlation with such success. Such data were by no means used as the sole criteria of selection, but, in so far as the items of the application blank were used at all, they could on this basis be assigned their relative importance and roughly combined to give certain total indications.

In all of these cases, and in general in the use of letters of application, the need is felt for more objective methods of securing information concerning the candidate's character and capacity. Thus an improved rating of applications which secures only 60 per cent of successful selections, leaves much to be desired, and should always be supplemented by whatever additional aids are available.

In the case of the studies of this traditional method of diagnosing character we have few objective facts with which to compare our judgments. It might seem reasonable to suggest that the consensus of opinion of all the judges would approximate such an objective standard. This,

however, cannot be assumed until it is demonstrated. It is entirely possible that some judge whose arrangements differ widely from the consensus of opinion is more nearly correct than is that consensus. Even in democratic countries the counting of votes does not necessarily determine truth.

In the following chapter we shall have occasion to consider judgments of character in circumstances where more reliable objective facts are at hand for use as a criterion of correctness. The letter of application often contains a photograph of the applicant. In many cases the candidates are requested to submit such photographs, and cases are even known in which it is announced beforehand that applications not accompanied by photographs will under no circumstances be considered. Presumably these photographs are supposed to reveal evidences of character not to be found in the letters, and since the photograph is usually the next thing to be considered, we may next turn to the judgment of photographs, as our second traditional method of diagnosing character.

CHAPTER III

HUMAN CHARACTER IN PHOTOGRAPHS

That character reveals itself in the features, and especially in the facial expression, is a belief of long standing and of quite general tenure. Even those who have long relinquished their faith in phrenology and physiognomy, with their attempts to tabulate the meaning of structural characteristics, believe that customary attitudes and feelings may be read from the total impression given by the face. A face or photograph, it is said, "looks intelligent," "beams with pride," "speaks defiance." Even the faces of the lower animals convey these meanings to us. Sometimes particular details are singled out as significant, but this is usually a rationalization, made after the general impression has already been given. Again the face is said to remind one of someone else who had given characteristics, as a result of which the trait is attributed to the stranger. But, aside from the arbitrary dogmas of particular professional "characterologists," there is no gen-

eral agreement on these details, either as between individuals or as between different photographs or faces.

Agreement of Judges

In spite of the disagreement on significant details, there is nevertheless considerable agreement between the verdicts which different people pass on the character denoted by photographs. A group of ten adults were asked individually to arrange a series of twenty photographs in an order of merit on the basis of several different traits of character. If there were no agreement among the judges, their arrangements would follow the laws of chance, and each photograph might be expected to appear equally often in any of the positions from one to twenty in the series. All the photographs would have about the same average position, which would be about rank number ten, and the average deviations of the verdicts of the various judges would be about five positions for each of the photographs, regardless of the trait in question. The following table shows the average deviations of the verdicts of the ten judges (averaging all the pictures together) for the seven different traits.

AVERAGE DEVIATIONS IN JUDGING PHOTOGRAPHS

Trait Judged	Average Deviation	Chance Deviation
Intelligence	2.86 places	About 5 places
Perseverance	3.32 "	" " "
Kindliness	3.55 "	" " "
Conceit	3.57 "	" " "
Courage	3.69 "	" " "
Humor	3.90 "	" " "
Deceitfulness	4.14 "	" " "

This means that in the long run one judge will place a photograph in a group of twenty, when arranged for a given trait, not over three or four steps away from the place to which others would assign it, instead of the five steps which chance arrangement would yield. This tendency toward more than chance agreement also varies with the trait. It is highest of all for intelligence and perseverance, lowest of all for humor and deceitfulness. and intermediate for the other traits here considered. In the case of humor and deceitfulness, indeed, the disagreement is almost as great as chance would produce, although in the case of intelligence the deviation is only a little more that half as great as would result from chance. These results suggest, then, that quite aside from the correctness of their opinions, people agree

fairly well on the facial expression that denotes intelligence, but that each has his own notion of the appearance of a deceitful face.

The practical question is of course not how closely different judges agree in their estimates of photographs, but whether these verdicts bear any relation whatever to the facts. Fortunately this investigation and several others similar to it, give us interesting information on this point. Is there "an art to read the mind's construction" in the photograph? The individuals whose photographs were used were ranked in order of merit for various traits by twenty-five people who were acquainted with all the individuals. These rankings were not based on photographs nor on features, but on actual experience with the people in question through a close acquaintance for a period in no case less than two years. Since human traits exist only in so far as they become manifest, we may fairly take the combined judgment of these twenty-five acquaintances to represent as correct a statement of the real characters as can conveniently be secured. To have been known by twenty-five associates for at least two years is, in a sense, equivalent to fifty years of acquaintance with one individual, and traits

that do not manifest themselves correctly in fifty experience-years are not likely ever to do so.

Unreliability of Individual Judgment

We can now compare the judgments of photographs, made by total strangers to the individuals represented, with the combined judgments of these twenty-five acquaintances. The comparison should tell us something fairly definite concerning the validity of judgments based on photographs by the method of general impression. If we represent a mere chance agreement, such as would be produced by shuffling the photographs repeatedly and averaging the rankings given after each shuffle, by zero (.00), we can by the appropriate statistical procedure, referred to in the preceding chapter, determine the degree of resemblance due to the correctness of the judges in their estimates. A resemblance of 1.00 would be perfect agreement between the ranking of the photographs and the combined verdict of the acquaintances. We might in this way secure coefficients of agreement lying anywhere along the scale from zero to perfection, that is from .00 to 1.00. We might even find negative relations, giving us such coefficients as —.24,

—.48, etc., or even— 1.00 in case one series should be absolutely the reverse of the other.

The following table shows what coefficients of agreement resulted from the comparison of the verdicts of the ten judges who arranged the photographs, with what we have described as the true or actual ranking of the people represented. The coefficients are given for all the judges in the case of three representative traits.

INDIVIDUAL CORRECTNESS OF JUDGES IN ESTIMATING COEFFICIENTS OF CORRELATION

Judge	Intelligence	Neatness	Sociability
I	.51	.11	.39
II	.11	.10	.08
III	.15	.29	.05
IV	—.27	.06	.49
V	.08	.24	.08
VI	.43	.41	.28
VII	.04	.11	.02
VIII	.39	—.09	.32
IX	.22	—.08	.00
X	.30	.02	.55
Medians ..	.19	.11	.18

The results are very instructive. The individuals judging the photographs tend slightly toward correctness rather than toward error, but the individual coefficients of agreement are so low

and so irregular as to be entirely useless. Verdicts of one judge, based on the photographs, are so inaccurate that on the average they give only a little more than chance agreement with the facts, and if one is going to rely on the verdict of a single judge, one might almost as well request him to shuffle the photographs and report which one comes out on top. Moreover, when one judge happens to yield a fairly high coefficient in judging one of the traits, he is entirely likely to fall very low in his judgment of other traits. And, as a matter of fact, in three cases in the table the individual judgments of the photographs gave negative coefficients with the true orders. Obviously, to accept the verdict of one judge would be unfair to the individual whose fate might be in question, as well as unfair to the judge compelled to assume this responsibility.

The Consensus of Opinion

But this does not necessarily mean that photographs tell us nothing about human character. As the table shows, the individual judges tend toward correctness rather than toward error. We might suppose that by combining the verdicts

of many judges, as we did in the case of the acquaintances, we might secure more accurate results, since the errors of one judge might tend to neutralize those of another, thus precipitating only their mutual correctness as the final result of the combination. That such a hope is in part justified is shown by the following table of results, which were secured by averaging the opinions of twenty-five women, in the one case, and twenty-five men in the other, when both were judging the character suggested by twenty-five photographs of women. Here as before we have compared the judgment of the photographs with the combined estimates of twenty-five acquaintances.

JUDGMENTS OF PHOTOGRAPHS COMPARED WITH JUDGMENTS MADE BY TWENTY-FIVE ACTUAL ACQUAINTANCES

Trait	Photos Judged by 25 Men	Photos Judged by 25 Women	Average of Men and Women
Neatness	.03	.07	.05
Conceit	.10	.27	.19
Sociability	.29	.29	.29
Humor	.21	.45	.33
Likeability	.30	.45	.38
Intelligence	.42	.61	.51
Refinement	.50	.52	.51
Beauty	.60	.49	.55
Snobbishness	.58	.53	.56
Vulgarity	.61	.69	.65
Medians	.36	.47	.45

Considering the average results it appears that by combining the verdicts of several judges the accuracy of the judgments based on photographs is increased. The individual judges intelligence with a correctness of only .19 on the average; the group judges intelligence with an average correctness of .51. In the case of sociability also the group correctness is greater. The individual, however, judged neatness, on the average, with a correctness of only .11, whereas the group judges it with no greater correctness. In the case of the other traits the correctness of the group judgment varies with the trait considered. Neatness, conceit, sociability, humor and likeability give very low coefficients even when the group judgments are used. But intelligence, refinement, beauty, snobbishness and vulgarity yield coefficients sufficiently high to be interesting and suggestive, although they are far from representing perfect accuracy.

In general then the investigations here considered suggest that photographs may be used to convey useful information concerning the character of the individuals they represent, if the proper technic is employed. The technic consists, first, in securing the combined estimates of many

general impressions, and second, in recognizing that such combined impressions give significant coefficients of correctness only in the case of certain traits. Many different investigations agree in finding that the traditional methods of diagnosing human aptitude and character give, in general, coefficients of correctness of not over .25. It is then useful to know that, at least in the case of certain traits, the method of the combined impression from photographs gives coefficients of correctness that are at least twice as great (.51 to .65 in the present instances) as those resulting from the traditional methods in their usual form.

Additional Evidence

Since this investigation was made and first reported, several other studies similar to it have been reported by other workers in this field. Pintner (Intelligence as Estimated from Photographs, *Psychological Review,* July, 1918) used rather dissimilar photographs of children whose intelligence he had measured by one of the standard intelligence scales, and had the photographs ranked for intelligence by various judges. Just as in our own case, the individual judgments tended

toward correctness, but the coefficients of correlation were small. They averaged only about .10 when individual judgments were considered, rising to .16 when the judgments of the total group of observers were combined into a group impression. In another case [7] the employees of a company were given standard intelligence tests, and their photographs were judged for intelligence by twelve people, who were unacquainted with the individuals represented. A special method was used which does not give figures precisely comparable with those reported above, but the indications are very much the same as those from other investigations. By the method used, the combined judgments give a correlation of .27 with the intelligence ratings as secured through the tests.

In the case of this traditional method, then, our results, though critical, are not entirely disparaging. The scientific study of the traditional methods is not undertaken with the hope of bringing them into disrepute. The intention is instead that of investigating, by controlled experiment, the degree of value attaching to these methods and the conditions under which greater value can be

[7] Anderson. "Estimating Intelligence by Means of Printed Photographs," *Journal of Applied Psychology*, June, 1921.

secured. Nor should our study be influenced by the practicability or impracticability of securing in actual affairs these more favorable conditions. We must first know what the favorable conditions are and in what respects the method in question yields reliable information. The continued use of the method then becomes a question of practical expediency.

CHAPTER IV

JUDGING ONE'S OWN CHARACTERISTICS

The letter of application has certain characteristics of its own, by virtue of its being a letter—its form and appearance, its vocabulary and general tone, its stationery, its choice of topics and their relative emphasis. The letter may also be accompanied by a photograph of the writer. Both of these features, the letter as a whole and the photograph, we have considered in detail. Special attention, however, must be given to a kind of material which may appear in some letters but not in others, and concerning which special provision is made in many formal application blanks. Distinguished from the narrative of fact and biography, we must recognize those estimates which the applicant gives of his own mental or moral traits, his aptitudes, virtues and interests—in general, his self-estimates.

In an application blank issued by a Public Employment Bureau, which advertises "Facilities for supplying high-grade skilled and unskilled

male and female help promptly," the applicant is advised, "Your answers to all questions will serve as a guide for the Placement Clerk." Three pages of questions are then presented, which the applicant is to answer, concerning himself. Many of these questions relate to matters of fact and biography, education, previous occupation, ancestry, health, etc. Such questions are not of immediate concern to us, although the question of their relevance or importance may fairly be investigated. But about half of the questions are of an altogether different type. In them the applicant is asked to give an estimate of his own characteristics, sometimes in vague and general terms, sometimes in terms of greater precision. The following questions, chosen from the many included in the blank, will represent the type:

Does your mind concentrate or skip around?
Have you self-confidence?
Have you patience?
Do you act impulsively?
Are you persevering?
Do you consider yourself absolutely honest?
Can you plan well, and carry out your plans?
Are you inclined to be lazy?
Are you sensitive?

Can you remember things well and for a considerable period?

Similarly, on a College Observation Chart, the student is asked to grade himself at various times during his course, by assigning himself scores of A, B, C, or D in each of various traits or abilities listed on the chart. The traits, many in number, are such as Imagination, Humor, Friendliness, Self-control, Poise, Spirituality, etc. In such a case as this the estimation of one's own traits may have a distinct value, quite aside from the accuracy or inaccuracy of the self-estimates. The student, in attempting to grade his inclinations and tendencies, may find his own characteristics clarified and noticed in the process. He may readily learn how imperfect is his knowledge of himself and be led to useful habits of self observation and criticism. In the former case, however, it is asserted that the self-estimates of the applicant are to be used in determining his destiny. Here the question of their accuracy or inaccuracy is not irrelevant. Our experiments on this point can be readily presented under three headings, each indicated by a specific question.

Accuracy of Self-estimates

How close to the truth does one come in attempting to grade his or her own characteristics? Twenty-five persons agreed to rate themselves in nine different traits, by indicating at what point they stood in the group of twenty-five, when these were arranged in an order of merit for each trait. Thus a self-estimate of 1 in Neatness meant that the individual believed herself to be the neatest person in the group of twenty-five, with all of whom she was personally acquainted. A score of 13 would mean that in such a series this person believed herself to belong in the middle position, etc.

All the members of this experimental group constructed such a serial arrangement of the group for each trait, each placing herself at that point in the series at which she judged herself to belong. The record was made in an apparently anonymous way, but unknown to the members of the group, accurate record had been kept of their judgments, by a secret method. Each person thus not only judged herself, but was in turn judged by twenty-four of her acquaintances. If we take the consensus of opinion (combined arrange-

ments) of the acquaintances as fairly representing the impression made by the individual on the world, this gives us a measure of her manifest character in the traits considered. We can therefore determine the accuracy of the self-estimates by comparing them with the consensus of acquaintances. The following table shows how many steps the acquaintances varied from each other, in judging an individual's traits, and also the average error made by the individuals in estimating their own position in the series.

ERRORS OF SELF-ESTIMATES COMPARED WITH ESTIMATES OF ASSOCIATES

Trait	Average Deviations of Judgments of 24 Acquaintances	Average Displacements of Self-Estimates from Position Assigned by Acquaintances
Neatness	4.5 steps	5.8 steps
Intelligence	3.7 "	6.0 "
Humor	4.5 "	7.3 "
Conceit	4.1 "	5.7 "
Beauty	3.8 "	6.0 "
Vulgarity	3.5 "	6.1 "
Snobbishness	4.8 "	5.1 "
Refinement	5.9 "	7.2 "
Sociability	4.7 "	5.4 "
Averages	4.4 steps	6.1 steps

If we bear in mind that purely chance arrangements of such a series, from time to time, would

give each individual the same average position in the series, and that the average deviation of all the separate positions from this average would be on the average a little over six steps, it is at once clear that these individual self-estimates are nearly as inaccurate as they possibly could be under the circumstances. The average displacements of the self-estimates from the positions assigned by acquaintances is 6.1 steps, approximately what it would have been had the names been shuffled instead of seriously and honestly arranged. And it is quite certain that the individual estimates in this investigation were rendered quite as seriously and as honestly as they would have been on a formal application blank. The inference seems to be that the Employment Bureau might just as well have tossed up a coin, heads meaning "concentrate," tails meaning "skip around," for example, as to ask the applicant to estimate his or her degree of possession of the traits in question. This of course is on the assumption that the questions are asked of the applicant as a method of eliciting information, not merely to discover what the individual will do when confronted by such futile questions.

Constant Errors in Self-estimation

Are self-estimates likely to err in one direction rather than another? Here we inquire not merely concerning the accuracy of self-estimates, but concerning any constant tendency or tendencies toward bias that may be present. In judging ourselves do we overestimate or underestimate? It may of course be suggested at once that our general knowledge of human nature tells us that individuals will overestimate their good points and underestimate their bad ones. But "general knowledge of human nature" is always to be mistrusted until it is verified under controlled conditions. It may, for example, be true that such a general tendency is present, but that it is more conspicuous in some traits than in others. And in some traits there may be no constant error at all. The experiment we have just described gives us material for investigating these tendencies, if we merely record whether the self-estimates place the individual higher or lower in the scale than the position assigned her by acquaintances. Representing a displacement toward the upper part of the scale and toward the lower part of it by (+) and (−), respectively,

and averaging algebraically the tendencies for all the individuals, the data are as shown in the following table, along with certain other information.

SHOWING CONSTANT TENDENCIES OR BIAS IN SELF-ESTIMATION
Data Secured from Study of Estimates of Fifty People

Trait	Constant Error	Number Over-estimating Themselves	Number Under-estimating Themselves
Refinement	+ 6.3	80 per cent	20 per cent
Humor	+ 5.2	78 " "	22 " "
Intelligence	+ 3.0	68 " "	32 " "
Sociability	+ 2.2	68 " "	32 " "
Neatness	+ 1.8	50 " "	50 " "
Beauty	+ 0.2	50 " "	50 " "
Conceit	—1.7	48 " "	52 " "
Snobbishness	—2.0	36 " "	64 " "
Vulgarity	—4.2	34 " "	66 " "

In general our expectations from "knowledge of human nature" are realized. Traits which we should on the whole characterize as "admirable" traits are overestimated; traits ordinarily classed as "reprehensible," on the other hand, are underestimated. And this occurs in spite of the fact that the conditions of such an experiment perhaps incline a participant to resist the natural inclinations to a greater or less degree, in so far as one might through modesty place himself at a lower

point than that at which he confidently believed himself to belong.

But this is not all of the story. Some traits, such as Neatness, Beauty, and Conceit, may readily be classified as "admirable" or "reprehensible," yet these traits show no considerable constant error of self-estimation in either direction. And the constant errors for some of the admirable and for some of the reprehensible traits are fully twice as great as those for other traits in the same general group. The very least that we can say, on the basis of the results, is that in discounting an individual's opinion of himself or herself, no "blanket allowance" can be made for all traits indiscriminately. The errors of self prejudice are greater in some traits than in others. It is possible that the degree of constant error measures the degree of desirability or undesirability attributed by the individual to the trait in question. The order in which the traits occur in our table would on this basis indicate their order of desirability on the part of our subjects, and the size of the error would reflect the degree of desirability or undesirability. Not having included those traits, we cannot infer that "Cleanliness is next to Godliness," but it may be safe to

judge that Humor is next to Refinement, in the evaluations of our judges. The fact that Sociability and Snobbishness, which are in a sense opposite traits, show opposite constant errors of almost equal amount lends a certain probability to this conjecture.

Evidence confirming these suggestions has recently been reported by Knight and Franzen.[1] These investigators compared estimates of the importance of traits for the judges themselves, estimates of their importance for the typical member of the group, and estimates for the ideal member of the group. There was a definite tendency "to place one's self nearer the ideal than the typical." "There is a higher association between what they believe they are and what they would like to be, than between what they believe they are and what their fellow students believe they are." In a related experiment students and professors showed "a clear tendency to speak well of themselves in those virtues considered of greater importance by them, and to rate themselves less highly in traits considered less vital." "We think that this tendency to over-rate one's

[1] F. B. Knight and R. H. Franzen. "Pitfalls in Rating Schemes," *Journal of Educational Psychology*, 1922.

self and the extent to which any one individual does it, has possible diagnostic value quite apart from the truth or falsity of the ratings themselves."

Qualification of Judges

Does the possession of a trait accompany ability to judge that trait correctly, either in self-estimation or in the judgment of others? Is the best inspector or superintendent of teachers one who herself excels in the art of instruction? Is the best literary critic one who is himself an artist in composition? Is exceptional journalistic aptitude a prerequisite of distinguished editorial work? Put in these forms the question assumes more than a theoretical interest, and it is a question on which practical policies do not seem entirely agreed. While we cannot pretend to solve the whole problem, we can at least show what was the case in our own inquiry. For we have measured the ability of each individual to judge himself and others, and we have also measured the standing of each judge in the traits considered. What relation exists between these various measures?

The table on the following page gives the

results. The figures give the correlations (coefficients of agreement) between possession of a trait, or standing in it, on the one hand, and on the other hand, ability to judge that trait correctly either in self-estimation or in judging others for their possession of it.

CORRELATIONS BETWEEN POSSESSION OF TRAITS AND JUDICIAL CAPACITY

Trait	Relation Between Possession of a Trait and Accuracy of Self-Estimation in That Trait	Relation Between Possession of a Trait and Ability to Judge It in Others
Humor	.87	.59
Refinement	.83	.38
Intelligence	.59	.49
Sociability	.47	.48
Neatness	.45	.22
Beauty	.15	.23
Conceit	—.22	.19
Snobbishness	—.27	.33
Vulgarity	—.37	—.24

In general the more "admirable" the trait, the closer is the relation between possession of it and ability to judge it. The three traits we have classified as "reprehensible" give either very low positive correlations or, in four cases out of the six, negative coefficients. The latter cases

suggest that lack of the trait in question is more likely to characterize those who are able to judge it, and this relation is especially clear in the case of judgments of the self. If then we define a trait as a desirable characteristic or aptitude, the inference is that in the long run the ability to judge the trait tends to be an accompaniment of the possession of it, and that the degree of the ability to judge varies directly with the degree of the possession of the trait.

Interesting confirmation of these general findings is reported by Allport and Allport from the Harvard Laboratory. Different people were asked to estimate their own intelligence by the rating scale method. They were later given intelligence tests, and their self rankings compared with their score in these tests. There was a definite tendency for those high in intelligence to underestimate their own ability, and for those low in this trait to overestimate themselves. The errors of the former group were also less than those of the latter, in magnitude. The correlation found between self-estimates of ability and scores in the Otis Group intelligence tests was —.67. Of the sixteen individuals in the group, only one of those who scored above average in the intelligence

measures overestimated his own ability, and but one of those who fell into the lower half of the group, as measured by the tests, failed to overestimate himself. The average error of the less intelligent half was more than twice as great as that of the more intelligent half.

These natural tendencies in self-estimation may be summarized briefly in the following way. The individual judges himself less accurately than others judge him, and on the whole self-estimates have only chance accuracy. The individual's judgment is moreover a biased one. He tends to overestimate or to underestimate himself according to the presumed desirableness or undesirableness of the trait. But those individuals who actually possess a given desirable trait in high degree are more accurate in their self-estimates for that trait than are those who possess the trait in lower degree.

Improved Technic

What suggestions by way of improved technic does the investigation of self-estimates afford? If we now refer back to a preceding page on which are given sample questions from the application

blank of the employment agency, it is clear that the traits considered in these questions are not all equally admirable, or reprehensible, as the case may be. Our results suggest that the correctness of the self-estimates of the applicant will vary with his conception of the desirability of the trait. An entirely practicable suggestion is that before the applicant testifies to the degree of his possession of the traits, he be asked to rate the traits considered, grading them according to their value or importance for the work in hand, or for the position for which he applies, or for life in general. Such a step in technic is easy of execution, and in addition to affording useful insight into the candidate's expressed conception of the value of qualities called for, it affords at least a suggestive criterion on the basis of which to check up his estimates of his own character. But perhaps the most important result of this inquiry is the definite evidence that self-estimates are misleading and that this traditional method of judging character lacks the accuracy that its prestige suggests.

CHAPTER V

THE PERSONAL INTERVIEW

We may assume that various stages in the conventional procedure of "sizing up" the applicant's character have now been completed. After the traditional manner the letter of application has been received and passed upon. The photograph has been examined and judged. The candidate's estimates of his own virtues and vices have been scrutinized. The traditional procedure has still further devices up its sleeve. It ordinarily requires the applicant to present himself for a personal interview, and very commonly instructs him to submit testimonials or letters of recommendation from others who know him. The interview and the testimonial are the two traditional methods not yet taken into account in our inquiry. For the present we shall be concerned, and that only briefly, with the interview.

Since standardized interview procedures have not been widely adopted, we can define the interview merely as an occasion on which, usually for a very short time, the applicant confronts some

manager, executive, foreman or employment specialist, who undertakes to pass further judgments on his character, and especially on his fitness for a particular place or type of activity. What happens in the interview we cannot specify, since these events vary with the interviewer, his purposes, and his general conception of what is significant. We may assume, however, that the applicant's appearance is inspected, his physique noted, his mannerisms, general bearing and deportment, his dress and the care of his person observed. He meets the examiner, and may be introduced to others in his presence. He is asked questions and given opportunity to express himself in speech and gesture. The import of the questions we cannot specify, but they are presumably relevant to his past, his plans, and his qualifications, his general habits, interests, and affiliations. In special cases he may be taken for a tour of the plant or institution, or he may be invited to luncheon, or otherwise observed in public. But the traditional interview does not involve these elaborations. It is more likely to consist of a brief inspection and conversation, lasting but a few minutes.

It is not our present purpose to attempt to pre-

scribe what should occur in an interview, to dictate questions that should be asked nor points that should be noted, nor to recommend any special mode of report for the interviewer. Such an undertaking can be usefully attempted only after a study of the particular requirements of a given case or type of cases. We shall instead concern ourselves with a study of the characteristic results of such interviews, from the point of view of their reliability. The method will be that of noting the amount of agreement between different interviewers, when they have passed judgment on the same candidates for the same positions.

Experimental Tests of Interviews

Fifty-seven applicants presented themselves for examination by a variety of methods. A small number were to be selected from these for appointment to positions involving personal salesmanship of a well-known service. In connection with this enterprise, which required a day and most of a night for its completion, twelve sales managers from prominent concerns agreed to interview the applicants individually and to rate them for their suitability for the positions in question. All of these managers, or judges, as we shall now call

them, were experienced in the judgment of character by this method, and a large part of their own activity consisted of such personnel selection. Each man was given a private room or compartment, allowed to interview each candidate for a definite period which was the same in all cases. Each judge was allowed to do anything he pleased with the applicant during his interview. He might talk to him, ask him questions, let him talk, put him through some exercise or test, try him out on any manner of proposition. But at the end of the interview each judge was required to assign to the applicant a designation, grade or description which would make it possible for him in the end to classify the candidates in several groups or to rank them in order of merit, avoiding ties so far as possible. The basis of the grading was to be "suitability for the position in question."

When all the interviews were finished, the grades were assembled and so tabulated and translated that each applicant received from each judge a rating in terms of his standing in the group of 57 applicants. Position 1 was considered the most suitable, position 57 the least suitable, and the intermediate positions indicated appropriate degrees of suitability. The complete

tabulation is too elaborate to be given here, but we may take a purely random sample as illustrating the nature of the results. Each applicant was given a number before the interviews began, these numbers being assigned in a random order. In the table on the following page are given the ranks assigned to the first ten applicants, by all of the twelve sales managers or judges.

The facts shown by the table are instructive enough. Almost any given applicant is likely to receive ratings placing him at any point in the scale, from first position to last. Applicant C, for example, is given position 1 by one judge, 57 by another, 2 by a third, and 53 by a fourth judge; in general he occupies positions all along the scale of "suitability." Much the same result is to be observed with all of the applicants. Occasionally an applicant is found whom all the judges tend to judge more or less favorably. Thus applicant I may be said to be favorably rated, on the whole, although even here the positions assigned him range all the way from 2 to 36. Some applicants again tend more or less uniformly to be rated low. Thus applicants A and B are on the whole rated low, but their ratings range from 6 to 55, and judge number twelve, who was unable to avoid

RANKS ASSIGNED APPLICANTS BY INTERVIEWERS

Applicant	Sales Managers											
	1	2	3	4	5	6	7	8	9	10	11	12
A	33	46	6	56	26	32	12	38	23	22	22	9
B	36	50	43	17	51	47	38	20	38	55	39	9
C	53	10	6	21	16	9	20	2	57	28	1	26
D	44	25	13	48	7	8	43	11	17	12	20	9
E	54	41	33	19	28	48	8	10	56	8	19	26
F	18	13	13	8	11	15	15	31	32	18	25	9
G	33	2	13	16	28	46	19	32	55	4	16	9
H	13	40	6	24	51	49	10	52	54	29	21	53
I	2	36	6	23	11	7	23	17	6	5	6	9
J	43	11	13	11	37	40	36	46	25	15	29	1

ties in his ratings, assigned both men to the same position.

When it is borne in mind that these judges were not casual people who were enlisted in the investigation, but expert sales managers, experienced interviewers and directors of personnel, and that the position (salesmanship) for which they were rating the applicants was precisely in the line of work in which they had developed expertness and acquired positions of responsibility, the inference is clear. However much the interview may be improved by better methods of inquiry and report, in its traditional form it is highly unreliable. No better evidence is required than the spectacle of two different expert interviewers, one rejecting an applicant as the most unsuitable of the group of fifty-seven, another selecting him as the choice specimen of the lot.

Similar Investigations

Binet once investigated the value of the interview method as employed by experienced teachers in estimating the intelligence of pupils. During an afternoon in the laboratory, three teachers interviewed independently five identical school children and reported upon their intelligence.

Each judge was allowed perfect freedom in choice of method, but the judgment was to be based on the results of the interview. Binet shows that there was scarcely any agreement among the various estimates, although each teacher had considerable confidence in her own opinions. He points out the lack of uniformity and consistency in the devices employed to discover the children's intelligence, and shows how unreliable and misleading such unstandardized procedures are. The replies to loosely formulated questions about school subjects, observations on current events, skill or expression in reading, memory, facial appearance, shape of head, behavior of the eyes, and similar details were cited by teachers as sure signs of intelligence. Some teachers were found who were sure that not more than once in a thousand times would their opinions be wrong. Binet also observed that these judges, before the close of the interview, were likely to utilize a rough approach to the method of tests and he uses the experiment as an occasion for pointing out the necessity for better formulated test methods and the close relation between the more exact methods of tests and the rough methods often relied on in the personal interview.

An array of data similar in character to the results shown in the foregoing table has been reported by Scott. Six managers in charge of the selection of salesmen in different districts, interviewed thirty-six applicants. "Each manager was instructed to assume that he alone stood between the applicant and the pay roll of the company. This was a responsibility that every manager was familiar with. Following the interview each manager made a report on each of the thirty-six applicants and indicated which was the most likely candidate, the second best, the third best, etc." Scott gives a detailed table showing the considerable disagreements of these expert interviewers. He concludes, "As a matter of fact, in the case of 28 of the applicants, these six managers disagreed as to whether the individual should be placed in the upper half of the group or in the lower half of the group. An inspection of the table shows much agreement among the six managers, but the disagreements are striking. Thus applicant 17 was thought to be the third best of the group by Manager C; but was placed thirtieth by Manager B. Applicant 18 was thought to be the best in the group by Manager E; but was ranked as tied for the thirty-second place by Man-

ager D. Yet there is reason to believe that these six gentlemen agreed even more closely than is the case with employment agents in general."

Improvements in Method

In recent years much thought has been given to the improvement of the interview as a method of judging character. Perhaps the most obvious suggestion afforded by our present inquiry is similar to that found in the case of the judgment of photographs. Since individual judges are prone to error, multiplication of the judges may well serve to eliminate the error and to set forth such truth as each interviewer is able to discern. The least that can be said is that the larger the number of interviewers, the more nearly does the final record approach to the impression the applicant is destined to make on the world at large, and this, we have seen, is by definition his true character. The practical adoption of the suggestion to increase the number of interviewers may of course encounter difficulties. Multiplying the number of judges involves longer time, greater cost, and more thorough record keeping. But errors in judging character also lead to delay, expense, waste, and extra effort. From the point of view of

an industry or an institution the question is largely one that can be solved by simple arithmetic, a direct computation of profit and loss by the two methods. From the point of view of society at large and especially from the point of view of the applicant whose character is passed upon and whose destiny is thereby definitely affected, the issue transcends the bounds of arithmetic.

Aside from the increased validity secured by the multiplication of interviewers, it is quite certain that the individual interviewer can increase the reliability of his judgments in most cases, by the proper attention to his methods. This does not mean necessarily reducing the whole interview to a rigid and formal interrogatory. Instead the informal impression, the vague general reaction, should be given its due weight, providing that the interviewer has learned that his reactions to personal appearance, attitude, bearing, manner, of others, fairly represent the reactions of others. Private "hunches," personal antipathies arising from particular idiosyncrasies such as complexion, dress, accent, diction, should not be permitted to overweight the judgment. But the general impression, which represents the most probable first effect of the presence of the candi-

date on others, is usually not unimportant, and is worth recording. It should, moreover, be recorded as an independent item, and not confused with the verdict based on other data.

Next to the importance of segregating personal impression from other items is the importance of discriminating choice of questions. The interview is so largely a matter of interrogation and reply that the question easily tends to become perfunctory, and to be merely a means of occupying the candidate while a general scrutiny permits the development of general impressions of favorableness or unfavorableness. The method of personal interview in many ways combines the use of application forms and the use of trade tests. In both these cases progress has come through sharp determination of the respective relevance of various questions and the appropriate weighting of the corresponding items in the report. In general, the interview should supplement the application blank and pave the way for definite and objective trade tests.

In the first place previous job analysis and specifications should enable the interviewer to know definitely what information he can look for that bears directly on the candidate's competence.

He should proceed directly toward this information, as soon at least as a few general introductory questions establish a general feeling of rapport between interviewer and candidate. Each relevant characteristic should be independently scored, in as objective a fashion as rating scales will permit, and, if time allows, definite record should be made of those acts, words, or indications which serve as the basis of the judgment. This will enable others to make their own judgments and to check up the impressions of the interviewer. Questions should be so framed as actually to elicit information, and not to permit of specious correctness through random replies or through suggested answers. In these respects the work of the interview is the same as that undertaken by the social worker, the physician, the lawyer, in the development of a personal history or a section of testimony. The art consists essentially in (a) establishing a favorable rapport; (b) discriminating between relevant and irrelevant questions and replies; (c) on the one hand the elimination of, and on the other hand the effective and deliberate use of, suggestive questions; (d) independent recording of facts elicited as distinguished from inferences drawn; (e)

standardization or consistency in form of verdict and in terminology of report; (f) sagacious synthesis of varied independent items; and (g) typical or representative personal reaction to that total ensemble of elements that constitutes the candidate's appearance and attitude.[2]

We have not of course assumed that everything hinges upon the outcome of the interview. The interview is only one of a series of traditional stages which we have set out to investigate as best we may. There remains still for our consideration that most dignified and widely patronized technic, the testimonial or recommendation, with the examination of which our next chapter will be occupied.

[2] Prospective interviewers may be referred to two chapters on "How to Ask Questions" and "The Observational Method," in the book on *Employment Psychology,* by Henry C. Link. These chapters are full of valuable suggestions looking toward the improvement of the method of the personal interview.

CHAPTER VI

RECOMMENDATIONS AND TESTIMONIALS

Letters of recommendation, in spite of the common addiction to them, are often sealed with a shrug and opened with a smile. These strange attitudes toward a hallowed institution have a complex origin. For one thing the motives back of the writing of such letters are not always unmixed. The letter may be only one way of "speeding the parting guest." The enthusiasm of the writer may indicate only his joy over a separation long overdue. More perfunctory comments, on the other hand, may be less closely related to the mediocrity of the candidate than to a personal apathy with regard to his destiny. Of the writing of testimonials there is no end, and the very familiarity of the task may conspire with the limitations of vocabulary to give a conventionalized tone to such letters.

With these and many other factors involved in the evaluation of testimonials we are not now concerned. Instead, we shall consider the judgments

of character passed by associates and acquaintances, when they are delivered under the relatively simple motivation of fairness, candor and accuracy. These opinions of associates, previous employers, supervisors, teachers, and others are presumed to be on the basis of a fuller perception of character than are those of the interviewer, the application clerk, and the inspector of photographs. Based, as they commonly are, on longer and more direct experience of the applicant's conduct, they represent better samples of the way the candidate's character is manifested. But this is far from implying that such verdicts are wholly reliable or invariably correct. What information has the experimental method yielded that may enable us to evaluate the validity of these judgments of associates?

Testimonial Disagreement

For one thing, the first table of results given in Chapter IV shows the average deviations of the judgments of twenty-four associates in reporting the relative standing of twenty-five people in several traits. These deviations (see page 49) range from 3.5 steps and 3.7 steps in judging Vulgarity and Intelligence to 5.9 steps in estimating

Refinement. Bearing in mind that chance arrangements would have given average deviations of only a little over 6 steps, the disagreement on Refinement is seen to be almost as great as chance would produce. The closest agreements, 3.5 and 3.7 steps, are over half as great as chance deviations would have been. Several conclusions are at once suggested.

Acquaintances disagree considerably in their estimates of the traits of candidates. In judging some traits they disagree almost as much as possible. Their disagreement varies considerably with the trait in question. It could also be shown that their disagreement varies with the candidate who is being judged. In order for there to be disagreement, some or perhaps all of the estimates must be in error. Since disagreement implies errors and errors are what destroy validity, we can profitably examine somewhat more closely into the nature and location of these disagreements.

Data are available from several studies in which longer arrays of traits were considered. The numbers of judges (acquaintances) were smaller and the results lack the finality that we could desire. But the principle involved is clear and the independent studies give such consistent

results that they must serve as the point of departure for future inquiry. In the one case (Cattell) twelve scientific men estimated the character of five of their colleagues, assigning them grades in all the traits. In the other case (Norsworthy) nine members of a college organization were similarly judged by five of their intimate associates.

In both cases the judges disagreed with one another in characteristic ways. In particular, their disagreement was measurably greater in some traits than in others. If in each case we take the average disagreement on all traits as a basis (100), we can express closer agreement and greater disagreement than this standard by figures correspondingly lower and higher than 100. This enables us to compare all the traits in both the studies, and to combine the two sets of results. In the following table the results are thus expressed. The average deviations of the judges have in each trait been divided by the average of all of the traits. Decimals have been dropped from all of the measures. The traits are arranged in order from most consistent to least consistent, on the basis of the combined results of both studies.

The Hierarchy of Consistency

The differences between the results from the two investigations are not great, and can usually be understood by reference to the personnel and circumstances of the two occasions. Most important are the average results. They disclose a fairly definite hierarchy of consistency, for the circumstances described. At the upper end of the series the traits are judged with approximately 80 per cent of the standard disagreement; at the lower end, with over 120 per cent of this standard. The traits fall into three not entirely arbitrary groups, which we may designate A, B, and C groups. The B group represents close to average disagreement, the figures being not more than 5 per cent removed from the standard (96 to 105 inclusive). The A group contains traits on which disagreement is relatively small; the C group contains those on which disagreement is relatively great.

If an associate or acquaintance reports upon an A trait, there is reason to expect the testimony of other associates to agree fairly well with this report. A single testimonial relating to such traits has relatively high validity. But if an associate

SHOWING THE AMOUNT OF DISAGREEMENT AMONG JUDGES IN ESTIMATING, ON THE BASIS OF ACQUAINTANCE, THE TRAITS OF OTHERS, IN TWO INVESTIGATIONS

Relative Divergence of the Various Judges

Trait	Cattell 12 Judges	Norsworthy 5 Judges	Average of Both	Classification
Efficiency	75	92	83	
Originality	95	77	86	
Perseverance	75	101	88	
Quickness	90	88	89	Class A
Judgment	100	78	89	Median 89
Clearness	104	75	90	Close Agreement
Energy	75	109	91	
Will	85	98	91	
Mental Balance	110	81	96	
Breadth	100	92	96	
Leadership	90	103	96	
Intensity	85	113	99	Class B
Reasonableness	115	86	100	Median 100
Independence	104	98	101	Fair Agreement
Refinement	90	116	103	
Physical Health	115	92	103	
Emotions	120	91	105	
Courage	100	119	109	
Unselfishness	115	106	110	
Integrity	104	130	117	Class C
Cooperativeness	125	113	119	Median 118
Cheerfulness	130	112	121	Poor Agreement
Kindliness	120	125	123	

reports upon a C trait, there is abundant reason to expect other associates to deliver radically different judgments. A testimonial relating to such traits should be viewed with caution, for it represents but one of many discordant estimates. Single testimonials relating to the B traits, while not inviting active suspicion, should nevertheless provoke a certain reserve. Some associates will agree with it fairly closely, but others will have opinions of their own, and all will be worth considering.

In general, then, the higher the trait stands in the hierarchy, the greater the validity of the single testimonial. The lower the trait stands in the table, the greater the justification for delaying action until all the votes are in. However incomplete this list of traits may be, it includes a fairly rounded analysis of character in the vocabulary of ordinary speech and letters of recommendation. Such a scale of validity constitutes a useful guide to the evaluation of the testimonial of character.

Moreover the analysis suggests a general principle, on the basis of which other terms or traits may be classified. The A traits we may designate as "objective," in the sense that they represent

reactions to objects and impersonal situations and tasks, and are likely to result in objective products such as inventions, factories, books, bank accounts, salaries, positions, records, etc. These objective products are definite manifestations of the traits in question and they are open to general inspection. The C traits, on the other hand, represent reactions to the presence and character of other persons. They are personal, social, moral; they do not so definitely produce objective products open to general inspection. Instead, they lead mainly to personal and emotional reactions on the part of others; hence we may designate them "subjective" traits. The B traits stand midway between these extremes or partake of both characteristics. New traits or terms may be inserted in the hierarchy with considerable confidence, under the guidance of this general principle.

A rough check on the validity of this table of traits is to be found in the data reported by Miner, from a study of the estimates of certain traits of students by their instructors. The designated traits were Common Sense, Energy, Initiative, Leadership, Reliability, and General Ability. In order to ascertain the reliability of these

estimates, Miner compared the verdicts of different judges in various ways, all of which gave much the same results. He does not call attention to the fact that some of these traits show higher reliability than do others, but his data show this to have been the case. If we take the results from his larger group of students, men in the School of Applied Science, the order of reliability, from high to low, is Energy, Leadership, General Ability, Reliability, Common Sense, Initiative.

Of these trait-terms, two are to be found in the table we have just been considering—Energy and Leadership. General Ability is perhaps nearest to our Efficiency, Reliability to Integrity, Initiative to Independence, and Common Sense to Reasonableness. If these interpretations of Miner's terms are fair, the table shortly to be given shows that his results approximate very closely to what our general table would have led us to expect. The data from the smaller group of women students, however, do not agree so well with our table. The three traits that have highest reliability in Miner's study of the men students and the three traits with lowest reliability, are grouped quite as they would have been inferred

from our table, although the precise order varies slightly.

Comparison of Two Experiments

Trait	Order of Reliability of Estimates	
	From Miner's Results	From Our Table
Energy (Energy)	1	2
Leadership (Leadership)	2	3
General Ability (Efficiency)	3	1
Reliability (Integrity)	4	6
Common Sense (Reasonableness)	5	4
Initiative (Independence)	6	5

Folsom has reported a statistical study of character estimation, in which college men rated one another for a series of traits. The traits selected were in only a few cases comparable to those in the table we have been considering, and the variability of judges was shown only by correlating the group judgments of two groups of fourteen judges each. In spite, however, of the different method, Cheerfulness and Kindness were found to be the least consistently judged of all the traits, and more objective characteristics such as Personal Appearance and Handsomeness were the most consistent. Perseverance, and Enthusiasm (similar to our

term Energy) were included in the list. These fall in the section of our table designated the A group, and Folsom also found them more consistently rated than Kindliness and Cheerfulness, which fall in our C Group.

Objective and Subjective Traits

It is obvious that the objectivity or subjectivity of these traits, in the sense in which we have used those terms, is not solely a function of the meaning of the trait. Depending on the circumstances of acquaintance, a given trait may tend to be either more or less objectively or subjectively displayed. Thus the estimates of employees by their employers, of pupils by their teachers, of children by parents, may be based on different sorts of data from those used in estimates of superiors by their subordinates, or of superiors or subordinates by each other. The precise manner in which the circumstances of acquaintance determine the validity and variability of character estimates represents a field of inquiry that is almost entirely unexplored.

That the "objectivity" of the trait, as measured by the agreement of judges, varies with the circumstances of acquaintance may be shown by the

following brief report of another experiment. Eight college students (juniors and seniors) rated five of their instructors for several traits contained in the Cattell-Norsworthy list. In the following tabulation, these seven traits are given and in appropriate columns the order of agreement of the judges, along with similar rankings from the data of Norsworthy and Cattell. The trait on which the judges agreed most closely is given first rank, the one on which they agreed the least is ranked last.

When the judges are coördinate in status with those who are judged, the results are in accord with the general table. Students judge students in much the same way that teachers judge teach-

Ranking of Traits, for Agreement of Judges

Trait	Teachers Judging Teachers (Cattell)	Students Judging Students Norsworthy)	Students Judging Teachers
Efficiency	1	2	5
Energy	2	4	3
Leadership	3	3	7
Independence	4	1	6
Cooperativeness	5	6	4
Cheerfulness	7	5	2
Kindliness	6	7	1

ers, from this point of view. But when the judges and the judged are not coordinate, different results may often be found. Thus the students agree most closely, in judging their instructors, on Cheerfulness and Kindliness, precisely those traits on which coördinates disagree most markedly. They disagree widely in judging such traits as Efficiency and Leadership. Obviously the circumstances of acquaintance, rather than the nature of the trait as such, are responsible for these differences. These students knew their instructors from actual classroom experience. When the instructor was cheerful, he was cheerful to the whole class, and his kindliness was not of the personal but rather of the institutional variety. By these traits the students were more or less similarly impressed. But the instructor's Leadership or Efficiency are not general classroom traits. Judgments on them would be considerably influenced by chance or variable individual knowledge of the instructor's extra classroom successes. As a result of these circumstances of acquaintance, therefore, the objectivity of the trait will vary.

Closely related to this general point is the result of comparisons reported by Mann, of esti-

mates of the traits of graduate engineering apprentices by their foremen and by experts who were their superior officers and less closely in touch with their actual work. "The order determined by the ratings by half the foremen agreed fairly well with the order determined by ratings by the other half (correlation coefficient .48); and the order of merit in the judgment of one expert agreed fairly well with the order according to the judgment of the other (correlation coefficient .53); but the foremen's order and the expert's order did not agree so well (correlation coefficient .24)."[1]

A recent investigation[2] of such problems in connection with the rating of teachers was undertaken at the Bureau of Reference, Research and Statistics, in the Department of Education, New York City. In this study teachers in six different schools were rated by supervisors and associates in ten different traits. The "objectivity" of these traits was determined by finding the closeness of agreement of the various judges on each trait.

[1] C. R. Mann. "A Study of Engineering Education," *Carnegie Foundation Bulletin*, No. 11, 1918.

[2] "The Reliability of Judgment of Personal Traits," by John Slawson, *Masters' Essay, Department of Psychology*, Columbia University, 1920.

This "objectivity" of the traits varied considerably from school to school. Of the eleven traits considered, "Punctuality" occupied second position in two schools and either ninth or tenth in the other four schools. "Professional Interest and Growth" occupied tenth place in one school and second place in another. "Understanding of Children" occupied first place in one school and tenth place in another. Four traits behaved somewhat more consistently, in that "they at least stay in either the upper or lower half of the series of eleven" in all the schools.

The investigator, Slawson, calls attention to the general importance of such a result, in the following words. "If this group disagreement is inherent in rating, then universal procedures become an impossibility. Rating will of necessity have to remain a local problem. The trait that lends itself best to objective valuation in one school of a large school system or in one department of a large industrial organization will not do so in another school or in another department. The choice of traits would then become a special group or special unit problem."

Detailed inquiry into the reasons for such discrepant results in various sets of judgments showed that the uniformity, explicitness and accessibility of criteria were the most important factors. General acquaintance was not found to be important, although "acquaintance with the activities appertaining to a particular trait" played a considerable rôle.

Such results clearly show that it is fallacious to lay down general statements of the precise relative validity attaching to judgments of acquaintances in the case of various traits. Instead, the safe procedure is that of deducing general principles, as we attempted to do in previous paragraphs, and of applying these principles to the particular situation in which recommendations are to be evaluated. If this process implies a certain expertness on the part of employment managers and executives, this is not to be lamented. Even in the case of more exact measures, such as trade tests and placement tests, it is the rule that tests found effective in the selection of workers for a given operation are not always effective in a different industry or in a similar industry under different managerial, technical or marketing cir-

cumstances. In the use of these methods also technical expertness is required. Expertness, in the sense of mastery of the scientific methods underlying the work, is rapidly coming to be the prerequisite of all effective forms of personnel work.

CHAPTER VII

COMMON SOURCES OF ERROR AND THEIR CORRECTION

Special Tendencies of Judgment

Certain common sources of error in the judgment of others, even under generally favorable circumstances, may be pointed out. The *central tendency of judgment* is an influence that has been clearly observed in judgments of all sorts, even in judgments of very simple perceptual materials. Judgments of time, weight, force, brightness, extent of movement, length, area, size of angles, have all shown the same tendency to gravitate toward a median magnitude, the result being that stimuli above that point in the objective scale are underestimated and stimuli below overestimated, while the median magnitude itself is invested with no constant error. Just as our experience with a class, race or social group results in the conception of a *type* which shall in some way represent the central tendency of the group, and from which the separate members shall

deviate the least, so also in an experiment on sensible discrimination we become adapted to the median value of the series, tend to expect it, to assimilate all other values toward it, and to greater or less degree to substitute it for them.

Such a tendency seems to be present in our judgments of human traits—we underestimate the brilliant and overestimate the stupid. Terman has shown very clearly the marked tendency on the part of teachers and parents to overestimate the intelligence of retarded children, and to underestimate the intelligence of superior children. He has moreover pointed out some of the factors underlying these errors. To what degree the central tendency of judgment affects the estimates of other traits than intelligence we cannot know precisely until we have more accurate means of measuring these traits objectively.

Another possible source of error is suggested by the fact that estimated traits correlate quite closely with one another. Thus in an earlier report of the estimates of college students by their friends the writer found the correlations of Intelligence with Neatness, Humor, Conceit, Beauty, Snobbishness and Refinement to be .39, .59, .44, .34, .43, and .49 respectively. In this connection

the following comment was then made. "How far these figures measure definite relations between different and specific traits, how far they measure the degree to which one's impressions of various traits conspire to make up one's notion of other characteristics, or how far they measure only the degree of confusion that exists as to the precise meanings of the various words, it is exceedingly difficult to say."

Thorndike has more recently reported similar findings. "In a study made in 1915 of employees of two large industrial corporations, it appeared that the estimates of the same man in a number of different traits such as intelligence, industry, technical skill, reliability, etc., etc., were very highly correlated and evenly correlated. It consequently appeared probable that those giving the ratings were unable to analyze out these different aspects of the person's nature and achievement and rate each in independence of the others. Their ratings were apparently affected by a marked tendency to think of the person in general as rather good or rather inferior and to color the judgments of the qualities by this general feeling." Similar tendencies were found to exist in the ratings of army officers, aviation cadets, and

school teachers. "Obviously a halo of general merit is extended to influence the rating for the special ability, or vice versa."

Kohs and Irle also call attention to this bias in the judgment of one trait by the impression of another. "It seems very probable that when one is passing a subjective judgment on the question of whether Person A possesses a certain amount of trait *a* or *b* or *c* or *d*, that his judgment of practically all these is affected by some constant factor *x*. For example here is Tom Jones. Bill Smith is requested to record a personal estimate of his character, habits, self-control, intelligence, sociability—whether excellent, good, fair, very poor. What probably occurs when Bill estimates, is that each of his judgments is affected by a constant factor, possibly unconscious, such as 'Tom Jones is an excellent fellow. I like him because his ideas are very attractive to me.' This example is not typical, of course, but is merely utilized to illustrate the point." College teachers who are required to grade the essays and papers of their own students are usually aware of the difficulty of separating the quality of the product from a bias for or against some special and often quite irrelevant feature of the student's person-

ality. The managers of political campaigns soon learn that trivial and irrelevant details of physique, taste or affiliation are likely to be as effective in determining the fate of their candidates as are actual fitness for office or the merits of platforms. Gowin found that successful executives were characterized by superior physique. It is quite probable that the relation here is not a biological but a psychological one—superior physique provokes a general confidence and trust, and suggests authority, power and responsibility. It may of course also contribute psychologically toward the self-confidence of the individual executive.

This "halo," in so far as it may be demonstrated, is not to be confused with what the writer has elsewhere called by the very awkward term "general standout-ishness." Thus mental tests of college students were found to correlate not only with estimated Intelligence (.62), but also quite closely with Humor (.55), Snobbishness (.53), Beauty (.40), Neatness (.36), and Refinement (.34). This may mean either "that a sense of humor, a tendency toward self-esteem, physical attractiveness and a gentle manner dispose one's associates to think favorably of one's general men-

tal endowment'' or else that ''an individual who has sufficient distinction to stand out prominently in any of these estimated traits is possessed of a nervous system which enables her to accomplish the work of these mental tests with corresponding efficiency,'' as well as to predispose the judges favorably toward her other traits.

The possible sources of error we have considered may be briefly summarized. There is first, the central tendency of judgment, which tends to deflect all estimates toward an average. In the second place there is the ''standout-ishness'' of individuals which deflects estimates of particular traits toward the general estimate of the personality as a whole. The third tendency is for the strong impression of one trait to bias the estimates of all others. The first of these three tendencies arises from a general feature of all judgments. The last two are in part justified by the general ''quality of organisms,'' on the basis of which many of the various traits of an individual do tend to greater or less degree to be generally high or low or mediocre. But partial justification in the long run is no excuse for overlooking the error in any particular case.

Testimonial Validity

Agreement of judges is of course not in itself a guaranty of the correctness of their verdicts. In the case of recommendations and testimonials, however, which are relied on in the absence of more objective data, the validity of an estimate is to a considerable extent a function of the agreement of that estimate with others. In the case of several traits, such, for example, as musical ability, trade skill in various fields, scholastic proficiency, and intelligence, it should be possible to check up the validity of such estimates by comparing them with actual measurement. On several occasions this has been done for the trait Intelligence, a trait on which judges are likely to agree relatively closely, in comparison with other traits. Thus the writer has elsewhere reported the correlation of the estimates of intelligence by associates with ability in intelligence tests. Lindsay has recently reported the correlation, with intelligence tests, of estimates of children's native ability given by teachers who had been with them in the classroom for not less than

one month. The results of these various studies are summarized in the following table:

	Showing correlation with intelligence measurements of:	
a.	Combined estimates, by twenty-four friends, of the intelligence of twenty-five college juniors	.70
b.	Combined estimates, by twenty-four friends, of the intelligence of twenty-five college seniors	.53
c.	Combined estimates, by five teachers, of the intelligence of nineteen children in a tenth grade history class	.52

Such correlations show that combined judgments do tend very definitely, but far from perfectly, to agree with the results disclosed by mental measurement. But too much should not be made of this point, since many tests have been selected as measures of intelligence by virtue of their correlation with such estimates.

Even if our examination of judgment tendencies should close at this point, the results would have an important practical bearing. But there is more to be said, especially with respect to the improvement of the traditional methods. One source of disagreement in the judgments of associates is in the indefiniteness or ambiguity of the terms used to designate the character traits. Different

judges, using the same term, may have in mind not entirely the same aspects of conduct. Again, different judges, using terms which to the reader are apparently quite unrelated, may have in mind much the same concrete conduct. The ambiguity of meanings is thus one source of disagreement or equivocation in testimonials.

Another source of testimonial invalidity is the lack of uniform standards and terms of report. "Very efficient" may mean in one letter "exceptionally able." In the language of another judge "very efficient" may mean only "satisfactory." "Clearness" may mean "rare lucidity" to one judge, and only "intelligibility" to another. Even pseudo-quantitative statements, such as "70 per cent intelligence" may mean "more than average competence" or may mean "feeble mindedness indicating institutional care," depending on the speech habits of the judge. Thus a grade of 70 per cent in school work is a fair passing mark. But an intelligence quotient of 70 per cent is suggestive of mental deficiency to the school psychologist. Descriptive adjectives and pseudo-quantitative scores are equally subversive of testimonial validity.

The Elimination of Variability

In the case histories of patients with nervous and mental complaints, undertaken by psychiatrists, the inadequacy of descriptive terms has often been realized. Here it is frequently necessary to delineate as clearly as possible the antecedent personality of the patient, and to record, as fully as may be, his character and conduct at the time of examination. Such case histories are often referred to by those not present at the examination, or are used for comparative purposes in determining the patient's progress, or in the comparative studies of individuals. In traditional forms these records tended easily to become lists of descriptive adjectives, expressing not the patient's conduct, but instead the examiner's interpretation or classification of his acts. Such terms as "depressed," "poor memory," "flighty attention," "excessive worry," "over activity," "excitement," "defective judgment," "inadequate perception," and so on, obviously do little toward fixing the picture for future reference. Nor do such terms convey to others any precise notion of the patient's behavior or condition.

In this field, therefore, it has been urged that

the examiner refrain from descriptive adjectives and record instead the actual things that the patient did, quote precisely his verbal statements, give the questions asked or tests administered with the actual replies or reactions given. This substitution of narrative and description for inference and personal interpretation tends to eliminate the errors of subjective opinion and to communicate the picture undistorted by the idiosyncrasies of particular beliefs, theories, or vocabularies. Until psychiatry can use methods of measurement instead of personal reports, it is in some such way that the errors due to ambiguous terms and variable standards must be minimized.

Various methods have been suggested for the elimination of these two sources of misunderstanding. Indefiniteness of trait-terms may be remedied, for example, by presenting, instead of the single term, a group of terms conceived to be related or synonymous or approximately equivalent. But this practice may invite only etymological doubts or disputes concerning the suggested equivalence and may otherwise distract the judge. A more useful procedure in practice has been found to be the narration of instances and con-

crete acts, or the inclusion of a brief explanatory sentence, stating more explicitly the meaning to be given the term.

Consistency in the terms of report has been sought in several ways. A simple method is to substitute for adjectives and per cent or letter grades, a statement of the position or rank of the individual among a group of his general class, in the trait in question. Conventionally a group of 100 persons serves as a background for the location of the individual. Thus instead of describing an employee's "integrity" (when this is clearly defined), as "poor" or as "D," or as "60 per cent" (implying in some minds, failure), it may be said of him that he is "10th in rank" among a hundred men of his class. This is intended to mean that only nine men in such a group of employees would have less "integrity," whereas ninety would excel him in that respect.

Such a characterization does not pretend to be quantitative; it does not presume to show "how much" integrity a man has, but rather "where he stands" in this trait. Justifiable as this technique is on psychological grounds, it does not work well in practice, in part perhaps because it does

violence to the more familiar use of numbers as signs of amounts. Perhaps the main difficulty, however, has been simply that of making practical use of any procedure before it is widely adopted and conventionalized.

Rating Scales

Attempts have also been made to introduce concrete scales or "man-to-man" scales for each trait, each step on the scale being represented by some actual person known to the judge. Thus the judge, before characterizing the candidate in "integrity" may be requested to construct a hierarchy of real people he has known. Mr. A is to be the most nearly perfect specimen of the trait that the judge has ever met. Mr. E is to be the acquaintance in whom the trait is most conspicuously weak. Mr. C is to stand half-way between A and E; Mr. B midway between A and C; and Mr. D midway between C and E. Thereupon Mr. X is to be graded for "integrity" by locating him on this concrete rating scale. In more elaborate forms, weighted numerical scores may be attached to these positions in various traits, according to the importance of the trait, in the attempt to provide for the summation of different traits into some total score.

Of course an independent scale must be constructed for each trait that is judged. One of these forms, devised by Prof. W. D. Scott, and developed by the Committee on Classification of Personnel of the U. S. Army, is now widely known as the "Officers Rating Scale."[1]

Such procedure undoubtedly has many good points. If faithfully executed it may produce more consistent ratings, both from different judges and in different cases from the same judge. But it presupposes a certain industry and a certain fidelity to the principle of the method, both of which actual judges seem inclined to avoid. Thus they tend to assign the numbers, letters, or other symbols directly, without actual production of and consultation with the various concrete scales which the method requires. What was in intent a very rough method of ranking thus degenerates into the more conventional and labor-saving use of mere descriptive terms or their symbolic equivalents.

1 A full description of the derivation and use of this rating scale is to be found in the second volume of *The Personnel System of the United States Army*, published by the Government Printing Office. A critical examination of the validity of ratings made with the aid of this and similar scales is reported by Rugg (*Journal of Educational Psychology*, November, 1921 to February, 1922).

A compromise between the demands of concrete rating and the laziness or hastiness of judges is to be found in various graphic devices. Thus a line drawn opposite a trait on the report sheet may indicate the stretch from the lowest to the highest specimen in the trait. By simply putting a mark on the line at the appropriate place the judge may indicate his estimate of the candidate's position in this series. More elaborate adaptations of this simple device may indicate the frequency with which the various degrees of the trait are likely to occur, thus roughly approaching the method of percentile ranking. The graphic rating device seems to provoke less personal resistance than do the other methods, and to require less explanation and time. If the report sheet is well organized, the graphic rating device appeals to many as an interesting indoor sport. It spares the tedious vocabulary of the conventional letter of appraisal, expedites correspondence, and pulls a higher percentage of returns. A sample report sheet based on these principles is here presented, to be used in making inquiry of previous employers or other acquaintances, concerning a candidate for employment as a domestic worker.

CONFIDENTIAL REPORT CONCERNING MISS A—— W——

Please rate the candidate named above, for the traits indicated in large type, from the point of view of her qualifications for service either as a general houseworker or in connection with special domestic duties.

Among all the individuals so employed, within the field of your acquaintance, where would this candidate rank if all were arranged in an order of merit for each trait considered? Indicate the position in each trait by placing a dot along the line, grading the candidate as accurately as you can.

Please sign your name here.....

	Lowest Fifth	Fourth Fifth	Middle Fifth	Second Fifth	Highest Fifth
SKILL IN COOKING Knowledge of processes, variety and tastiness of products.					
SKILL IN SERVING MEALS Knowledge of customs, dexterity, promptness and speed.					
SKILL IN LAUNDRY WORK Strength and quickness, quality and carefulness of work.					
NEATNESS AND CLEANLINESS As to personal appearance, care of premises and utensils.					

	Lowest Fifth	Fourth Fifth	Middle Fifth	Second Fifth	Highest Fifth
GENERAL HEALTH Robustness, freedom from incapacitating illness and physical handicaps.					
ENERGY AND EFFICIENCY Confidence in handling work, promptness and system, reliability in emergencies.					
PERSONAL INTEGRITY Honesty, truthfulness, sobriety, general level of moral conduct.					
TEMPERAMENT Cheerfulness, courtesy, coöperativeness, freedom from temper, irritability, loquacity.					
LOYALTY Fidelity to duty, care of employer's interests, interest and pride in work.					

Specific instances, characteristic acts, and similar observations will be appreciated, and may be noted on the back of ths sheet.

Mail to W—— B—— (Envelope enclosed.)

A Model Inquiry Form

The advantages of such a form are obvious, once it is prepared. Only strictly relevant traits are included, these being known from a previous analysis of the traits of successful workers. Each trait is briefly but clearly defined. The number of traits included is small, and the graphic method of rating, aided by the suggested division of all such candidates into five groups according to merit, makes it a very simple matter to make a systematic record of the judge's impressions. Three groups of traits are distinguished. The first three on the form belong to the "objective" group; the last three belong to the social or "subjective" group; the middle three traits stand midway between these objective and subjective traits. The A, B, and C traits may thus be independently evaluated, although no indication is given to the judge that his reports on some traits are considered less reliable than his reports on others. Opportunity is given to report specific incidents, acts, and observations.

Suggested Improvements

We have, then, numerous improvements to suggest in connection with the traditional appraisal by associates. When a general and informal letter is called for, it should be suggested that the letter contain, where possible, concrete accounts of acts or attitudes, rather than merely an array of descriptive adjectives. It is, however, easier for the initiated to write such letters than it is to secure them from others, who may neither appreciate the principles involved nor find the time and convenience for adherence to them. In general, therefore, it is better to provide the reporter with a standard report sheet or testimonial form. On this form should occur a limited number of definitely relevant trait-terms, succinctly defined or illustrated. The traits should be selected with reference to their position in the local hierarchy of validity. Objective, ambiguous, and subjective traits (the A, B, and C groups of the preceding sections) may all be included, but they should be segregated for readier evaluation of the returns.

A graphic mode of record or some similar mechanical or checking scheme should be used, which eliminates the necessity of verbal ingenuity and

decisions of choice on the part of the judge, who can thus devote his whole attention to the task of appraisal. Enclosed with a stamped and addressed return envelope, such an invitation to judge one's fellow men is calculated to promote both the peace of the reporter and the validity of the report. But the validity of the ratings is subject to many qualifications, even under the best of intentions; for this reason it is desirable, wherever possible, to secure a record of the actual facts on the basis of which the judgment is passed, as well as the statement of the verdict itself. Since most of our judgments of associates are based on the summation of numerous minor impressions, it is only in special instances that the concrete data of such judgments will be available.

The peace of the reporter is only an irrelevant gain unless the validity of his report is sufficient to give it reliable accuracy. That such reliability is not attained, even under the most favorable conditions that are practicably possible, is shown by Rugg's investigation of the accuracy of various rating scale methods. After a study of many thousands of ratings, by various methods, especially in educational and military institutions, this investigator concludes that, for such purposes

"The point cannot be made too emphatically that we should discard these loose methods of rating once and for all. We cannot justify wasting the time of our school administrators and deluding our teachers with fictitious 'ratings' and 'marks.' Even on one of the so-called 'standardized' point rating schemes a *single* rating has little or no scientific validity." The only circumstances in which Rugg reports such ratings to be sufficiently accurate for practical uses in education, for example, involve the following rigorous requirements. The final rating used must be the average of at least three independent ratings, made by individuals thoroughly acquainted with the candidate, on comparable and equivalent scales as objective in character as those of the man-to-man comparison variety, when these have been formulated in conferences under the instruction of one skilled in such technique. The impossibility of even approaching such conditions in most practical situations is of course obvious. We may, therefore, profitably consider, as our next topic, the possibility of developing more objective methods of identifying and measuring human traits.

CHAPTER VIII

MEASURING MENTAL COMPETENCE

Even when trait terms have been clearly defined, there is no guaranty that a list of such terms represents distinctive character elements. "Integrity" for example may be a complex product or resultant of the organization of several features of the mental structure, just as water is a product of elements which might combine with each other or with other elements in numerous diverse ways. Character terms, such as those of daily life, represent effects produced rather than psychological functions or elements of personality. Scientific attempts to secure more accurate pictures of the mental pattern itself proceed by some method of mental measurement. Measurement begins by devising tests. A test is a standard situation which is calculated to provoke a response that can be quantitatively or qualitatively evaluated. In a sense the writing of an application and the ordeal of the interview are tests. But they occur under conditions that are

not controlled and yield results not readily capable of accurate evaluation.

The test with which measurement begins is preferably a very restricted task or situation which can be readily repeated, controlled, scored, and evaluated. Measurement is facilitated if the task involves but few types of material, few processes or functions or variables, definite and identifiable capacities, interests, feelings. Progress consists in establishing, for such identifiable details of mental make-up, significant and specific tests. Any conceivable task, performance or bit of conduct may be made to constitute a test, if it goes through the necessary and somewhat complicated preliminary stages. Until these and various subsequent stages have been completed, the task does not constitute a measure of competence or character. To describe some of these stages in the development of tasks into tests and tests into measures, for the appraisal of human traits, is the purpose of the present chapter.

Tests and Measurements

Any situation or performance may become the material of a test. Sharpening a lead pencil, defining words, threading a needle, speed of read-

ing, adjusting a carburetor, filing a saw, accuracy of spelling, appreciation of jokes, hitting a target, preference of pictures, memorizing a poem, drawing a sketch, milking a cow, detecting logical fallacies, are more or less miscellaneous and trivial examples. As enumerated, they are merely feats, reactions, feelings. They become tests when standard materials are prescribed, standardized technique adopted, standard instructions formulated and used, standard scoring methods devised. The tests become instruments of measurement when, in addition to the satisfaction of these requirements, the scores can be translated in terms of an authentic background, scale, or context of scores, and interpreted as symptoms of present status or signs of future conduct.

For example, for the feat "filing a saw" to become a test and then to evolve into a measure, first (materials) the size, condition and quality of saw, file and vise must be specified. Next (technic) the position of the worker, the illumination, and other details of procedure and setting must be indicated. Third (instructions) the same directions must be given to each subject and these directions should be understood. Fourth (scoring) the effectiveness or quality of the act must

be stated in prescribed terms and the units that go to make up the score must be agreed upon.

Suppose all these conditions to have been observed, and an unpracticed candidate found to be able to file the standard saw, with the standard tools, by the standard method, under the standard directions, with a standard quality of execution, in thirty minutes. Thirty minutes is then the score in the test. But it is a "raw score," and in itself relatively meaningless, hence not a measure of anything except the actual duration of the act itself. But suppose that by extended use of this saw-filing test it is found that unpracticed subjects, who do this feat with this score, are those who, having a moderate native interest in tools, are sufficiently dexterous to attain the status of a journeyman carpenter and no more. The test score can now be translated into a measure, in terms of skill level attainable, of the subject's "carpentering ability." This is the essential process involved in all scientific character analysis. A test is found which serves as a symptom of more extended functions or prospects than are involved in the task itself. The inference from the detail to the larger pattern is thus more

than a mere judgment, in so far as it rests on a foundation of established data.

Principles of Construction

The test act itself may be planned or standardized on the basis of one of several principles. In the case just described a *Standard Task* was set, the completion of which is necessary before the score (time required) can be determined. Another method uses *Homogeneous Material* as its basis. The score is then the number of items or task units acceptably accomplished in a fixed time. Thus in the saw-filing test, the number of teeth filed in ten minutes might constitute the score. The method of *Graded Tasks* is also often employed. Tasks of the same nature, arranged in a series of steps of increasing difficulty, comprise such a scale. The score is then the distance along the scale which the subject's competence will carry him, either in a fixed time or without time limit. A method of *Miscellaneous Gradations* also presents a series scaled in difficulty, the steps, however, being represented by varied kinds of tasks or materials. Another common principle is based on *Response Values*. The stimuli or situa-

tions are the same for all candidates, but achievement is measured in terms of the quality of the reactions made.

Readers familiar with the names of common tests now widely used will recognize the Form Board and the Puzzle Box as cases of the method of Standard Task, the Substitution and Cancellation tests as examples of Homogeneous Materials, the Stenquist Construction test and most Vocabulary tests as instances of Graded Tasks. Miscellaneous Gradations are illustrated by the Binet-Simon scale and its various adaptations and revisions. The principle of Response Values is used in the Kent-Rosanoff test for Community of Ideas.

Principles of Expression

Along with the standard modes of choosing acts, tasks, or situations, should also be considered the various standard methods of expressing the measure of achievement or response. Five chief methods are in common use. In some cases the raw scores or *Original Units* are used in expressing the result, as in ordinary athletic scores and in the determination of "critical scores," failure to attain which means rejection. Or these raw scores may be translated into *Percentile Units,*

in which the standing or rank in a representative group of one hundred candidates becomes the final expression. Related to this is the use of *Distribution Units,* in which the raw scores are divided by some standard measure of the variability of the representative group in the feat in question. Such a measure shows the individual's position with respect to the average of the representative group. The principle of *Developmental Units,* such as age units, translates the raw scores into terms of the average age of immature individuals who attain this as their characteristic score. Such a measure states the "mental age" or the "maturity" of the candidate in the trait in question. The use of skill levels also resembles that of Developmental Units. Finally, *Absolute Units* may be devised, as when the accomplishment is shown to be a given number of equally perceptible steps superior to a zero accomplishment.

Original Units have little place in mental measurement. Until placed in a context of other scores they have little or no meaning, and they cannot under most circumstances be combined nor directly compared with scores from other tests. Thus measures of height and weight, in terms of inches and pounds, can neither be compared nor

combined significantly. But if expressed in Percentile Units, Distribution Units, or Developmental Units, such comparisons and combinations are possible. By such methods, for example, it may be discovered that a given individual is "more superior" in one trait than in another. Measures of many traits in such comparable terms make possible the graphic portrayal of the "profile" or "psychograph" of the candidate in the array of tests or traits considered.

Illustrative Test Procedures

The following selections of tests that are easily described will serve to illustrate concretely these various methods of standardizing materials and of expressing results:

Word Building.—Consider the letters A, E, I, R, L, P. You can make words of those letters. Thus you can say E-A-R spells a word. But you cannot make R-A-T because there is no T there, nor can you make P-I-L-L for there is only one L. Make as many words as you can and write them down. Do not use any letter that is not there, and do not use the same letter more than once in the same word. (The score is the number of words you can make in five minutes, counting all words correctly spelled, excluding obsolete and foreign words, and abbreviations. About 60 words are possible.)

If we know how many words can be made by average children of various ages we can state the raw score in terms of Developmental Units. Thus a score of 13 words means a mental age of 15 years in this test, since the average fifteen-year-old can make 13 words in the time allowed. Or, if the candidate is a college freshman, and three minutes are allowed instead of five, a score of 13 words locates the candidate as one probable error (1 P.E.) below the average college freshman. The average freshman score is 16 words and the probable error (representing the variability of 100 freshmen about this average) is 3 words. Further, we can say that a candidate standing at this point in the distribution of ability is at the 25-percentile, or is number 25 or at rank 25 in a representative group of 100 freshmen, when they are arranged in order of competence in this test, from poorest to best. In this test, then, we have illustrated the three principal methods of expressing the measure (Developmental, Distribution, and Percentile Units).

We have still to illustrate the other methods of determining the materials. In a certain sense we can use the foregoing test to illustrate the method of Homogeneous Materials. If we consider that

each of the words in the complete list represents an equally valuable item, as is actually the case in assigning the score, these words then represent homogeneous items which the candidate has to "think of." The number of these homogeneous tasks he can "handle" in the time set, thus constitutes his score. A clear example is afforded by the following section of a familiar "substitution" test. The key at the top indicates a number as belonging to each of the various geometrical forms. The task consists in writing in each of the forms below, with pencil, the number which belongs to each, taking each form in turn across the page, and consulting the key as often as necessary. Here the various forms and the task connected with them, constitute homogeneous items in the task. The raw score may be translated, as in the case of the word building test, into terms of Developmental Units, Distribution Units, or Percentile Rank.

The method of Standard Task is conveniently illustrated by the following assignment. There is another common English word, and only one, that can be made of the letters contained in the word CHESTY. What is that other word? Here the task is unequivocally specified. The response

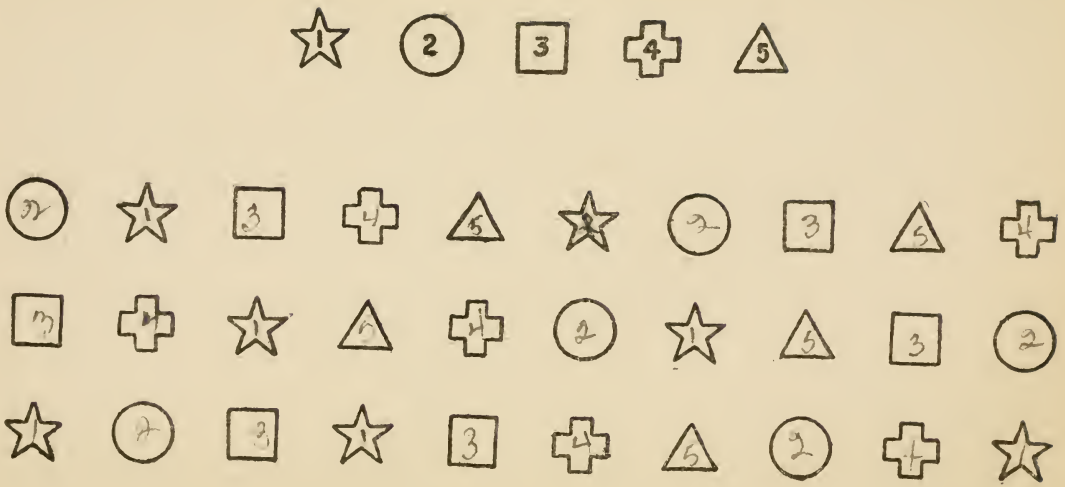


FIG. 1. Substitution Test.

will be of the "all or none" type, since degrees of success are not possible, either one succeeds or fails. The score will be the time required to discover the correct reply, or else it will be simply "failure."

The method of Graded Tasks is neatly illustrated in a rough form by the following "re-arrangement" test. Each set of letters, when properly re-arranged, spells the name of a familiar animal. The tasks are similar in each case, but their relative difficulty has been determined by trial on numbers of adults, and the words have been arranged in order of difficulty. This is a

REARRANGE THE LETTERS TO MAKE NAMES OF ANIMALS

1. snkuk	7. bnsio	13. hapletne	19. rotligala
2. niol	8. leum	14. kmyeon	20. laglroi
3. rede	9. seomu	15. duonh	21. tinram
4. tgrei	10. shroe	16. ycteoc	22. lettru
5. barze	11. goroauka	17. lqriurse	23. clreoicdo
6. selwae	12. delopra	18. kunmpchi	24. peatnole

"rough" illustration, because it is not known whether or not the steps from word to word represent equal steps in "difficulty."

For exact purposes it is desirable that the steps

in such a graded scale be equal. But equality may be present in one or more of various senses. The advance may represent a step in felt difficulty, a step in the frequency of solution, in the average age of attainment, in the time required to accomplish it, etc. In any given case the nature of the criterion of advance must be taken into account. The following test illustrates a more accurately graded series of tasks which are similar in general character. The steps are equal in the sense that the average time required for each item in the scale, by educated adults, is in each case except the first item approximately twenty-five per cent (25 per cent) more than the time required for the item just preceding it in the scale. The average time required for each item by a group of college students, expressed in fifth-seconds, and the average deviations, are given in the following description of the test:

Analogies Test

A pair of words is given, the second of which bears a certain relation to the first. Then a third word is given in a separate column. You are to give a fourth word, which stands in the same relation to the third

word, as that in which the second word stands with respect to the first word. Any word that satisfies the requirement is accepted as correct.

First Pair		Second Pair		Av. Time in fifth sec.	A.D.
eagle	bird	shark	fish	8	1.5
multiplication.	division	addition ...	subt	10	2.5
opera	hear	movie	see	12	2.8
boat	water	train	rain	15	3.0
gallops	horse	bites	dog	19	7.0
color	blue	tool		24	8.3
cart	horse	automobile .	ring	29	17.9
eyes	face	lake		37	19.0
courage	cowardice	virtue		47	25.5
hand	fist	nation		60	32.5

The Method of Miscellaneous Gradations is well represented by the following section of the Stanford Revision of the Binet-Simon intelligence scale. At each age level (Developmental Unit) is placed a set of tasks, which although they vary considerably in their apparent character, are alike in the sense that the ability to handle them in the prescribed fashion and with the prescribed degree of success is achieved by the typical child of that age or more, but not by the typical child of lower ages. The tasks are miscellaneous but gradation is nevertheless present.

Year III

1. "Show me your": nose..... eyes..... mouth..... hair..... (3 of 4).
2. Names key..... penny..... closed knife..... watch.....pencil..... (3 of 5).
3. Three objects in one picture: Dutch Home Canoe Post Office
4. Gives sex.
5. Gives last name.
6. Repeats (1 of 3): (*a*) I have a little dog; (*b*) The dog runs after the cat; (*c*) In summer the sun is hot.

Year IV

1. Compares lines (3 of 3, or 5 of 6).
2. Discriminates (7 of 10): Circle..... square..... triangle..... other errors.....
3. Counts 4 pennies (no error).
4. Copies square (pencil, 1 of 3): *a*..... *b*..... *c*.....
5. Comprehends (2 of 3): What must you do when you are

 (*a*) sleepy. (*b*) cold. (*c*) hungry.
6. Repeats (1 of 3): 4739..... 2854..... 7261.....

Year V

1. Compares weights (2 of 3): 3—15..... 15—3..... 3—15.....
2. Colors (no error): red..... yellow..... blue..... green.....

3. Aesthetic comparison (no error): *a*..... *b*..... *c*.....

4. Definitions (use or better, 4 of 6):

chair	doll
horse	pencil
fork	table

5. Patience (2 of 3), 1 minute each): *a*..... *b*..... *c*.....

6. Three commissions. Key on chair..... brings box..... shuts door.....

The only method not illustrated in the foregoing account is that of Response Values. In its more common form the value of the response is determined by its possession of some quality. The following example will serve to illustrate this method. The test is one of "free association." The words in the first column serve as "stimuli." They are read, one at a time, and the candidate is requested to respond by the first single word that "comes to his mind" upon hearing the stimulus word. Examination of 1,000 people shows that for each stimulus word some responses are "preferred;" that is, they are given by large numbers of people. Responses to a given word may be graded in their frequency, commonness, or banality, by indicating the number or per cent of people who give them. The candidate may, for each word, be given a score which is the per cent of peo-

ple who give the same response that he does. The average of his scores or some other index of their tendency, will give a measure of his "community of ideas." The quality of the response, or its value, here consists in the degree to which it represents the responses of others. In the table the words in the first column are the stimulus words. In the other three columns are given the three most frequent responses, and the number of people who give them.

SAMPLE WORDS AND PARTIAL RESULTS FROM THE KENT-ROSANOFF FREE ASSOCIATION TEST

Stimulus	Three Most Frequent Responses, and Their Absolute Frequency in the Test of 1,000 People		
table	chair (267)	wood (76)	furniture (75)
dark	light (427)	night (221)	black (76)
music	piano (180)	sound (95)	song (68)
sickness	health (142)	death (115)	doctor (62)
man	woman (394)	male (99)	boy (44)
deep	shallow (180)	water (134)	ocean (93)
soft	hard (365)	pillow (53)	easy (34)
eating	food (170)	drinking (166)	bread (46)
mountain	high (246)	hill (184)	valley (90)
house	home (103)	building (78)	barn (74)

The method of Absolute Units may be readily illustrated by one of the scales now in common use for measuring the results of education. Thus

the quality of the candidate's handwriting, from the point of view of its general "excellence" may be measured. A sample of his usual handwriting is secured. As the result of preliminary investigation, a handwriting scale is ready, consisting of specimens of the penmanship of different people. These specimens are arranged as a scale, of such a sort that the first specimen has just zero value, inasmuch as it cannot possibly be read, although in its general appearance it suggests an attempt at writing. The succeeding specimens of the scale represent equal distances along the scale of "excellence," in the sense that these advances are equally perceptible or, as is more usual, agreed upon by equal numbers of judges, that is, equally often perceptible. The candidate's handwriting may be placed alongside such a scale, at the point where its quality is judged to equal the excellence of the scale specimens. The score given it is then the number of "perception steps" by which it excels the zero specimen. Samples of such scales can be found in many of the manuals on educational measurement. They are too elaborate to reproduce in this brief survey and their valid use requires expert knowledge.

Primary Meaning of Test Records

At this point occasion may be taken to point out just what information it is that a test yields. Primarily and directly, it affords a measure only of that which it actually measures. Thus, when a specimen of handwriting is measured on the handwriting scale, there is afforded only a measure of that particular specimen of handwriting, not of the candidate's handwriting ability. If the specimen was written in excitement, haste or fear, it may indicate the tendency of his penmanship under such circumstances. If it was prepared at a given speed, the measure may suggest the candidate's general tendency when writing at that speed. If it was prepared on a moving train, its significance is again limited. The point of these comments is in general that the immediate measure is only a measure of the immediate product. Two specimens of writing may be thus compared without any special precautions, from the point of view of their excellence as present products. But unless all other conditions of performance, such as rate of writing, controlling purpose, emotional condition, degree of training, and so on, are

kept constant or otherwise accounted for, the inference cannot be extended from the actual specimen to the ability of the candidate.

Similarly, ability to re-arrange the letters to make the names of animals constitutes, directly, only a measure of that actual process. It is an error to describe the test as one of imagination, invention, intelligence, proof-reading ability, interest in verbal matters, or anything else, until, by a further process of inquiry this particular test has been empirically found not only to measure what it does directly measure, but also to indicate whatever more general facts it may be found to signify. Furthermore it is not even a direct measure of the process itself unless various other factors, such as time allowed, instructions given, motivation, and other variables are specified, kept constant or otherwise accounted for.[1]

Choice of Tests

The manner in which a test is constructed and in which its results are capable of expression may often determine its suitability for a given purpose. Thus the method of Standard Task is less suit-

[1] *Cf.* A. I. Gates, "Experimental Study of Reading and Reading Tests," *Journal of Educational Psychology*, September, October and November, 1921.

able for group procedure than are the methods in which a fixed time limit is established for all candidates. The method of Graded Tasks is likely to cover a shorter range of capacity than some of the other methods, inasmuch as a type of process or material sufficiently difficult for the higher degrees of ability may present no tasks which can be accomplished by the lower degrees. Or conversely, materials and processes suitable for the lower degrees may not readily afford tasks sufficiently difficult to enable the measurement of the more able. A more or less specific trait, such as ability to spell, may best be measured by the method of Graded Tasks, whereas for the measurement of more general and vaguely defined traits, such as mechanical skill or intelligence, Miscellaneous Gradations may be more readily prepared. Other illustrations might be given, but these should suffice to suggest that both the construction and adoption of tests for definite purposes call for a background of information and a comprehension of technique which the novice is not likely to possess.

The method of expressing the results may also condition the uses to which a test may be put. The principle of Developmental Units, illustrated by

the use of mental ages, has the advantage of picturesqueness and concreteness, because of the well-established age concept in other connections. But it has many limitations and it may readily lead the novice into absurdities. Thus it can only be applied in the measurement of the immature and of the inferior adult, since standards based on the principle of mental age commonly cease at points ranging from eleven to sixteen years. Moreover it is only by a kind of convention or on the basis of some analogy that the steps from age to age can be considered units. Developmental units, in other words, are not likely to be units in any other sense. To have the capacity of twelve year olds in one test may mean to be as good as the average adult. In another test twelve year ability may mean marked deficiency. Therefore psychographic charts, such as those given in the last chapter of this book, based on the principle of Developmental Units, can lend themselves readily to misinterpretation.

The uses of Percentile Units and Distribution Units have many advantages. In the first place they can scarcely be used without being understood, and this is a great gain. Further, they fit in with the demands of various statistical pro-

cedures, without which mental measurement cannot progress. Although such units were among the first employed in mental measurement their adoption has been relatively slow. They do not make a strong appeal to popular fancy and their use requires a certain intelligence and technical knowledge. These methods also require somewhat more elaborate preliminary data and norms than does, for example, the method of Developmental Units. The ability of a stupid adult may be assigned a mental age equivalent, if norms on children only are available. But he could not be given a percentile or distribution rating unless norms on representative adults were also available.

Since nearly all children go to school and since tests can be put to good use in their instruction, it is relatively easy to secure representative norms up to about twelve years of age. Beyond that point it grows increasingly more difficult to secure representative data on the basis of which percentile or distribution ratings can be assigned to the individual chosen at random. These various points in connection with the principles of construction and expression are only samples of the many considerations required in the safe and profitable application of the methods of mental meas-

urement. The reader who cannot think of several others is only on the threshold of a complex field of technical study. Any novice can use a yard stick, but, in spite of a fairly widespread misconception to the contrary, there are no yard sticks for the measurement of human character.

Individual and Group Methods

From the point of view of economy and speed the mental examination of candidates assembled in groups has many advantages. On the other hand greater accuracy and flexibility, increased reliability and fuller knowledge result from the independent examination of each individual. Many tests in common use are not suitable for group procedure since, for example, they may involve somewhat complicated technique, oral responses, elimination of error through refusal to accept inadequate or incorrect replies, scoring in terms of time, or various other complications. Limitations on the part of the candidates also detract from the reliability of test scores secured by the group method. Although success or high score in group tests may usually be taken as an indication of competence, failure or low score may result from one or more of a number of factors.

Thus illiteracy, sensory defect, physical impediment, illness, misunderstood instructions, in the case of those otherwise competent, may produce inferior performance. In any given circumstance it is useful to know the degree to which group survey scores agree with results from individual examinations.

Thus in one case, in which the interest lay in the measurement of intelligence, both group survey methods and individual examination were employed on a group of about one hundred candidates. The differences in mental age as determined by the two methods were then computed, and the average difference was found to be 1.3 years. Of all the cases, 48 per cent showed a difference of less than one year. On the other hand, in 10 per cent of the cases the difference was three years or more. In 79 per cent of the cases the score by the group method was lower than that from the individual examination, and on the average the group scores were about one year inferior to those based upon individual examination. In only three instances did individuals score lower in the individual examination by as much as two years.

On the other hand, a group of forty-seven adults

who had failed to make scorable records in the group test were individually examined, and it was found that 75 per cent of these adults were below 8.5 years of mental age. For the most part, then, those who completely failed in the group examination were actually mental defectives. In general such results, which have also been confirmed by other investigators, suggest that group methods give fairly accurate measures of the average competence of the group as a whole, and fairly reliable scores for the competent or the superior individuals in the group. But low scores in group examinations cannot be taken with assurance to imply incompetence, in the absence of other evidence.

The reader will perceive that no attempt has been made to familiarize him with the many, many hundreds of tests of human aptitudes and characteristics now employed by psychologists. To master all these, to know their history, their detailed character and significance, their technic, their uses and applications, and their implications in the analysis of human character, has in recent years become a professional occupation, similar to surgery, dentistry, law, and engineering. The materials grow daily more elaborate and the field

more complex. The purpose of this chapter has been, through a general survey of the main principles, to convey a rudimentary notion of the nature of the methods of construction that have emerged in the development of this field up to the present time.

CHAPTER IX

GENERAL COMPETENCE AND SPECIAL APTITUDE

Specialized Traits

The foregoing chapter was mainly concerned with the description and illustration of some of the fundamental concepts and technics underlying the development of tests and measures. The problem of *interpreting* such measures and of determining what they show about the candidate's character is in a sense an independent enterprise. The final score may be given merely as a measure of precisely that thing which was done. Usually, however, it is hoped that this act will serve as a significant index of more comprehensive aspects of the candidate's mental equipment or character.

When tests of achievement are employed they fall more or less clearly into two groups. If rating in the tests is compared with other measures or estimates, some tests are found that appear to be positively and relatively closely

correlated with all or with many kinds of proficiency. Thus a test such as that of completing sentences from which certain words have been omitted, has been found to correlate positively with many other tests, with estimates of associates for "general ability," or "alertness," with teachers' judgments of mental capacity, with success in school work, with age up to mental maturity, and with proficiency or status in economic, occupational or general affairs. Such a test appears to be a useful index of some general quality of the organism. It serves as a significant symptom of general competence. In the literature of mental measurement the term "intelligence" is commonly used to designate this general quality of the organism in so far as it is manifested in proficient action and competent conduct.

Other tests are found (such as pitch discrimination, color sensibility, strength of forearm) which do not show this general correlation with proficiency. They may be more or less unrelated to other performances, responses, or attitudes, in which case they represent or suggest special aptitudes, traits which are distinguishable from general competence. The functions indicated may be limited to the precise acts involved in the test, or

they may be more extensive than this and point to aptitudes or tendencies in a certain direction or with certain kinds of material, situation, attitude, or action. Moreover, although distinguishable, identifiable, and measurable, their actual manifestation in the individual's unanalyzed life may vary with interest in them, with opportunity, and with general competence. Illustrations of such special inclinations, aptitudes, or interests may perhaps be found in such characteristics as eloquence, visual imagery, skill in literary composition, wit and humor, leadership, mechanical dexterity, aesthetic sensitivity, musical ability.

In any attempt to use tests for the purposes of vocational guidance or personnel selection, it is important to know whether the measures reveal general competence or special aptitude, and whether the measures reflect native capacity solely, or are also determined by exercise; whether the contemplated occupational activity calls for special aptitude, skills or interests, or whether it requires only such general understanding and dexterity as is involved in intelligence. Since both general competence and special aptitudes may characterize different individuals in varying degrees or amounts, it is also important

to know what level of capacity, in either or both respects, the work demands, and whether the capacities required must be natively present or whether and to what degree they are susceptible of acquisition or cultivation through training and practice.

An Experimental Demonstration

In the following table are brought together the results of several different investigations, in which standing in various tests has been correlated with measures of occupational proficiency or with estimates of supervisors. The figures show, for each test, the correlation in the case in question, with occupational ratings. Although the precise results and their final interpretation must be held subject to many qualifications, the data adequately illustrate a common finding in this field.

In so far as these compilations of different investigations, conducted under the same general supervision, afford comparable data, they suggest that Directions, Color Naming and Opposites, all of them familiar laboratory tests, tend to reveal general competence. They give definite positive correlations with output and with esti-

CORRELATION OF CERTAIN TESTS WITH OCCUPATIONAL RATINGS, UNDER PARTICULAR CIRCUMSTANCES

Test	Stenography	Typewriting	Hand Sewing	Machine Sewing	Teleph. Operat.	Routing Clerk	Secretarial Work	Sales Girls
Directions	.46	.13	.27	.12	.70	.83	.54	.14
Color Naming.	.34	.43	.43	.36	.38	.00	.38	.21
Opposites	.45	.14	.41	.44	.00	.98	.40	.55
Analogies	.31	.00	.33	.58	.00	.71	.43	.32
Checking	.00	.40	.26	.00	.00	.82	.22	.13

mates of supervisors in most of the eight different types of work. Analogies and Checking, equally familiar tests, are not such clear instances of general indices. Analogies, to be sure, fails to correlate with only two of the occupational activities. But Checking gives zero correlation with three types of work, and correlations below .26 in three other cases. In only one field (Routing Clerks) does this test correlate more closely than .40 with occupational proficiency, under the circumstances here described.

The tentative value of these particular correlations should not require further emphasis. Other studies with the same tasks, with different workers, and under different local, technical and managerial circumstances, might very well yield

different figures. And if the measures of occupational proficiency had been more accurate than they were, there is reason to believe that the correlations would have been closer in the case of some of the tests than the figures indicate. Correlations are impaired both by inaccuracy in the tests and by inaccuracy in the occupational ratings, and one of the difficult problems in establishing the meaning of tests in this field is that of securing accurate objective ratings of individuals already engaged with varying degrees of success in the occupations concerned.

But the general principle suggested in these results is a fairly universal one. Some aptitudes are more highly specialized, restricted in scope, and more nearly elementary in character; other aptitudes manifest more general functions, attitudes, interests, and qualities. The general qualities are measured by attested intelligence tests. The special aptitudes are measured by tests particularly adapted to that specific purpose. Since the equipment of special aptitudes sets limits to the repertoire of intelligence or general achievement, and since general competence determines the appropriate and effective use of special aptitudes, both types of measurement are necessary in

scientific analysis of character from the point of view of competence or proficiency.

Temperamental Qualities

A third group of tests may show significant relations to traits of character not suggested either by "general competence" nor by "special aptitude." Instead they may disclose more or less general aspects of personality for which the terms competence and aptitude are not entirely appropriate. Traits commonly denoted by such terms as patience, honesty, cooperativeness, kindliness, sympathy, stubbornness, sociability, will serve as suggestive examples, although little is yet known about the conditions of these traits. In a preceding chapter we have shown that it is on these traits that associates differ most widely in their estimates. This suggests that many of these qualities or traits may be not so much traits of one individual as traits of a pair of people. A's cooperativeness undoubtedly depends on the personality of B with whom he is cooperating. Should the resulting degree of cooperativeness be attributed to A or to B? Apparently it should be applied to A and B as a team, rather than to either alone. Is a man's cheerfulness a trait of the man

or a function of the circumstances in which he is placed? We expect a stupid man to be quite generally stupid, and a musical person to be musical under all ordinary circumstances. Obviously cooperativeness and cheerfulness do not manifest themselves in this way, and such traits must be more closely related to circumstances and to social combinations of individuals than are the specific aptitudes and general competence.

It is in the judgment of this third group of characteristics that the traditional methods find their greatest difficulties. Similarly, it is in the measurement of these traits that the methods of mental measurement have made the least progress. A more detailed discussion of these difficulties will be presented in a later chapter, when a survey is given of the present status of such measurements. For the present we are chiefly concerned in drawing the distinction between three actual or possible groups of tests, those which relate to the general quality of the organism's mental activity, those which relate to special aptitudes and restricted proficiencies, and those which may be shown to be chiefly related to the temperamental characteristics, in so far as these are found to be individual traits.

Illustrative Cases

Illustrations of tests designed to measure general mental competence, alertness, or intelligence are easily found. Among the simpler processes that seem to serve this purpose are the various completion tests, in which verbal or pictorial materials are used. Trabue's Language Scales and Healy's Picture Completion Tests are typical instances. Tests involving the understanding of instructions, the following of directions, the perception of logical relations, the learning of codes and the association of symbols, solution of problems, detection of similarities and differences, rapid and accurate classification of materials, and similar acts or processes have been found to afford good indications of general mental competence. Such tasks are included in most of the systems of intelligence measurement, such as those of the Army Alpha, the original Binet-Simon, the Stanford Revision, the National Intelligence Tests, the Otis Group Tests, the Performance Tests standardized by Pintner, the Thorndike College Entrance Tests, the Association Tests of Woodworth and Wells, and the various other

sets and systems assembled or devised by numerous workers in the field of mental measurement.

Definite tests for the measurement of special aptitudes, in so far as these are not closely dependent on general competence, are no less numerous. Seashore's tests of musical talent, which afford measures of sense of pitch, sense of time, sense of intensity, sense of consonance and tonal memory, seem to relate to particular aptitudes of this character. Stenquist's Assembling Tests involve a type of mechanical insight and manual dexterity which, although to a certain degree dependent on experience, do not correlate closely with achievement in intelligence tests. Various educational tests and the trade tests measure special acquisitions and skills in which general competence, however much it may be involved, is not directly revealed by the test scores. Among the educational tests are such as are designed to measure knowledge or ability in reading, arithmetic, penmanship, composition, languages, algebra, spelling, and other academic subjects. Among the trade tests are the various oral, picture and performance tests which identify

the grade of skill of a carpenter, plumber, mechanic, chauffeur, telephone operator, boiler-maker, cobbler, etc., as being that of a novice, apprentice, journeyman, or expert. The numerous tests of sensory and motor capacities, such as those for auditory and visual acuity, color vision, tactual and kinaesthetic discrimination, simple reaction time, speed of movement, strength, steadiness and coördination, are designed to measure functions in which general competence is not closely concerned.

Among the measures of the more temperamental characteristics may be listed the various "interest" tests, of which the Mechanical Interest Test developed by Rice is the most definite example. In a later chapter will be given an account of some of the methods proposed for the measurement of such temperamental traits as will-pattern, emotion and mood, nervous stability, suggestibility, aggessiveness and eccentricity of thought.

In so far as technic of construction and expression of results are concerned, all the methods described in the foregoing chapter may be utilized in all three of these groups of tests and measures.

The present distinctions are not in technic but in the interpretation of results. And it is in the interpretation of results that the test method, having developed adequate measures, becomes a method of judging character.

CHAPTER X

THE METHOD OF THE SAMPLE

Practical enterprise is often content with measures of the candidate's net ability to effect certain results, regardless of the particular pattern of traits brought to bear on the task. The psychologist on the other hand, in so far as he is under the theoretical rather than the practical motivation, continues to be interested in the more detailed features of performance. So indeed must practical enterprise, also, in so far as it undertakes the selection and training of apprentices or novices and in so far as it is interested in the candidate's future progress rather than in his immediate performance. In a previous study of the methods of vocational psychology [1] the distinctions were made among four types of specialized vocational tests, instituted mainly in the interests of practical enterprise, but also used to contribute toward psychological analysis and individual guidance. These are the methods of the *Miniature,* the

[1] *Vocational Psychology,* D. Appleton and Company, New York.

Analogy, the *Sample* and the *Empirical Procedure* or *Correlation.* In the subsequent developments of this field the Miniature and the Analogy have justified the earlier predictions—they have not contributed materially either to knowledge or to practice. The method of the Sample and that of the Empirical Correlation have, however, undergone material development. In this chapter and the one to follow, some of these more recent accomplishments are to be considered.

The Procedure of Sampling

The method of the Sample proceeds by requiring the candidate to meet definite specimen situations, representing the tasks which his character and capacity will be expected to handle in the actual job. In terms of his success with these samples, his character and capacity are rated in the lump, as of such a sort as to fit him to do the specified work or to meet the specified situations with a given degree of success. No analysis of traits is made and ordinarily no distinction is made between native capacity and acquired skill. Although the method itself is entirely non-psychological, it has appeared that the psychological attitude and familiarity with the statistical

technic and the experimental procedures associated with psychology, are necessary to put the method on a solid foundation. Some of the most definite advances in the measurement of skill and knowledge, notably the various educational scales and the trade tests, have been due in large measure to the energetic and critical work of psychologists.

In the earlier account of the method of the Sample, it was described as that of taking an actual piece of the work to be performed and sampling the candidate's ability by his success in this trial. Thus in connection with the recommendation of clerks and assistants from among the boys in commercial high schools it is common to test their ability from time to time through their course by assigning them small pieces of work similar to that which they might later be required to perform in business offices and shops. Finding addresses and numbers in a telephone directory, carrying out involved verbal instructions and directions from memory, computing calculations, recommending action on the basis of their figures, making out a trial balance, a trial chemical analysis, are common forms of this type of test. In certain cases such specimens of work

have been devised in or taken into the psychological laboratory and the worker watched more closely and measured more exactly.

It was also pointed out that for the vocational test of this type to be at all significant, either the sort of work involved in the occupation must be fairly uniform and homogeneous in all its different circumstances (as in the case of typewriting at dictation, or in the work of filing clerks, accountants, etc.) or else there must be included a large number of samples representing all the various unrelated sorts of work.

Trade Test Methods

In the later development of the method of the Sample, as in the trade tests and educational tests, progress came mainly through the development of technic for the selection, not so much of complete samples as of significant ones. Once the selection was accomplished the work of establishing the most probable skill level of the candidate as that of novice, apprentice, journeyman, expert, or as fourth grade, fifth grade, sixth grade ability, was fairly straightforward. The method of the sample, then, does not proceed by analysis of the particular trait-pattern of the candidate. Instead

it measures his success in handling a significant task by whatever array of natural capacities, interests, education, and practice he may be able to bring to bear on the task. It represents, in a sense, a judgment of character, but of character in the lump, and in terms of present status only. It reveals neither the origin of the skill manifested, nor does it, taken alone, delineate the prospects of future achievement. Since detailed accounts of the derivation of such product scales and trade tests are available in the appropriate reports and manuals, we need give here but a brief sketch of them, as representing the modern development of the method of the Sample.

The trade tests afford good illustrations of this method. Here the samples chosen are either of actual trade skill or of trade information. Thus a truck driver is measured in trade skill by having him make a sample trip, accompanied by the judge. But the tests of different men, in different circumstances, and ratings by different judges, are given objective character and definition by their adhering to the general principles of a mental test. A standard outfit, standard trip, specified situations and emergencies, prescribed methods of scoring and of interpreting the score,

give results that have a validity far exceeding that of the mere subjective opinion of an inspector. The "performance test" for general blacksmiths neatly illustrates the method of the Sample, inasmuch as only one of hundreds of possible tasks is used. But the task is so chosen as to have demonstrated value in differentiating the various skill levels recognizable in the trade. It is thus a significant sample and serves as an index of the total equipment of information and skill possessed by the candidate. In this test standard equipment, materials and tools are provided, standard instructions formulated, and a standard scoring plan prescribed. The candidate, presented with a blue print specification for the making of a twisted hook of definite shape and size, carries out to the best of his ability the various processes of preparing, welding, twisting, punching, and bending the materials provided, so as to make a product conforming to the specifications. When such hooks are made under these conditions, are scored according to the prescribed criteria of excellence, and the standard scoring units are assigned, a score of less than two points characterizes the novice, whereas blacksmiths' apprentices may score anywhere from 2 to 131

points. Journeyman blacksmiths are found to score from 132 to 143 points, inclusive, and only experts at the trade are likely to score above 143 points. This performance trade test thus not only illustrates the method of the sample but also the construction method of Response Values and the expression principle of Skill Levels or Developmental Units.

The performance test, as a method of sampling, approaches fairly closely to the traditional practice of hiring on trial. It differs from this method in its emphasis on standardization. In other forms of the trade test, the picture method and the oral method, a more indirect judgment replaces the fairly direct perception of the candidate's skill. In these forms actual operative skill is not sampled. Instead, sample measurements of the information concerning the trade processes and implements or materials are used as indices both of the total stock of knowledge and of the most probable degree of skill thereby implied. Through careful preliminary investigation, standard oral questions and pictures calling for explanation, naming of parts, etc., are selected. These are so chosen that the score is not subject to the judgment of the examiner nor to chance replies, but

can be objectively determined from the answers given by any one who can read the scoring directions. They are chosen, moreover, so that the total scores from a given array of questions indicates, in the light of previously ascertained norms, the most probable skill level of the candidate. A series of ten to twenty questions bearing solely on trade information which is elicited through oral interrogation or the use of pictures, can be selected so as to afford a significant and practicable index of actual occupational competence at the date of examination. This fact in itself shows the very valuable contribution to the analysis and measurement of human capacity made by the method of the Sample.

Educational Measurement

Such sample measures, when properly organized and applied, afford something more than information to be used in the selection of workers for immediate assignment and occupation. They may be used also to measure the progress under training and the effectiveness of instruction, if adequate precautions are used. In the light of a candidate's previous experience they may also be used as signs of his probable future attainments.

It was as a matter of fact to the confluence of two active enterprises—the measurement of educational products in school administration and the search for improved technic in trade interviews, that the striking advance of trade test methods, in the military emergency, is to be attributed.

In our earlier references to the measurements of the quality of hand writing, as an example of the use of Absolute Units, the use of the sampling method in educational practice is also illustrated. The educational product scales classify the candidate's present attainment in the subject to which they relate. Ability in penmanship, composition, drawing, history, arithmetic, algebra, Latin, grammar, reading, and various other academic skills, constituting as each does "a combination of a complex set of intellectual and muscular coördinations," are constituent elements of a candidate's total character. In practically all these instances of educational measurement the method of the Sample is used as a basis for judgments of the more complex patterns of information and dexterity. Either sample tasks confront the pupil whose general rating depends on his success with these samples, or else a sample of his own work is rated on a concrete specimen scale.

The utility and validity of the method, whether in industry or in education, rests on the possibility of securing, in a limited set of test samples, reactions which, though far from exhausting the complex field of facts and actions, are nevertheless representative or significant of the whole structure.

It is quite beyond our present purpose in this discussion to enter into the details of the construction, use, and interpretation of trade tests or of educational product scales. Each of these is in itself an elaborate field, and they have been amply set forth in other places. Our purpose has been simply that of exemplifying, through them, the very striking advance in the diagnosis of human capacity in the case of one of the methods, that of the Sample, which though early recognized and roughly employed, as in the traditional trade interview and the conventional academic examination, awaited the application of psychological and statistical technic and the coöperation of schools and industries with scientific men, before it developed into anything like a method of diagnosis.

CHAPTER XI

CHARACTER ANALYSIS THROUGH CORRELATION

In the earlier account of the methods of vocational psychology á method of analysis and appraisal was pointed out which is especially characterized by the small number of assumptions which it involves. It is further characterized by the considerable amount of labor and of expert understanding for which it calls. This is the method of Empirical Procedure or the method of Correlation. Beginning with a group of individuals who differ in known degrees with respect to some trait or capacity, this method seeks for clues to those differences. A great number of measured details, such for example as physical features, education, individual interests, ability in various single tests, etc., are compared with the known status of the individuals in the trait in question. Each measured feature is independently correlated with the known status in the hope of

finding some, among many, that are significantly and consistently related to the trait itself.

By such empirical procedure certain tests may be found which are positively correlated with the more general trait or capacity and hence may serve as symptoms or signs of it. It is not pretended that the significant tests are miniatures of the larger pattern, nor samples of its operation, nor even that they involve the same fundamental functions. The test records show significant correlations with the ability or trait, so that excellence, mediocrity or inferiority in the one is as a mere matter of fact, empirically determined, indicative of similar standing in the other. Hence with an assurance depending on the degree of correlation, standing in the one may be taken as a symptom of standing in the other. This is not an unique logical procedure. In the same way a patient's temperature, or the activity of his reflexes, or the analysis of his blood or urine may serve as a sign of the more general bodily condition.

Obviously, several symptoms are more reliable than any single one of them, and in the correlation method several significant tests are sought for.

Preferably those tests are finally chosen all of which are closely correlated with the trait or ability but none of which is closely correlated with the others. In such a team or battery of tests each reveals, according to the magnitude of its correlation, a significant feature or part of the more general pattern, and no one test tends merely to duplicate the contribution of another. The greater the number of tests, the higher the correlation of each with the established criterion or trait or ability, and the lower the intercorrelations of the tests themselves, the more completely will the team reveal the total character pattern of the individual. By the proper statistical technic it is possible not only to correlate the tests with the trait and with one another, but also to determine the degree to which each reveals information not already contributed by the others.

Empirical Correlation Procedure

Certain prerequisites of this method must be enumerated. In the first place there must be an experimental group of individuals differing in known degrees or standing in known relations to each other, with respect to the trait or capacity

concerned. If the trait is an acquired pattern, such as trade skill, the tests may then serve as signs of the presence of this skill, but they do not, on this basis alone, serve to predict the degree of ability to acquire such patterns. Ability to do the tests may have been acquired along with skill in the trade and hence may be part of or a result of that acquisition.

The use of this method for predictive purposes, as in vocational guidance, in the adoption of children, or in the selection of candidates for instruction or development, requires that the tests be made on unpracticed or naïve individuals, who shall then proceed through training and exercise to develop such degrees of the trait or skill pattern as they can. Comparison of original test scores, made before such practice, with skill ratings after practice, will then yield correlations which, in varying degree will indicate the prognostic value of the tests.

In both the diagnostic and the prognostic, that is the selective and the predictive use of this procedure, correct trait or skill ratings on the experimental group must be available. To secure ratings which will be actually objective, or at least

reliable, is by no means easy. It is in fact one of the most formidable of the various problems encountered in the use of the method. Even when individuals have long records of achievement behind them, so that evaluation rests more directly on perception than on judgment, measurement of ability is far from simple. For one thing the circumstances and setting of the various achievements are usually neither alike, nor constant, nor obvious. Production records, sales sheets, and similar data, for example, do not entirely reflect the relative values of employees. They do not include the contribution made by each toward the "good will" of the public nor the prospect of continued or future returns.

Nor are such data in themselves necessarily comparable, since different workers may have encountered different geographical conditions, different competition, commodity differences, differences in consumer resistances. Salaries and wages are equally subject to irregular influences, such as length of service, for example. Task and bonus systems of rating inevitably involve a degree of arbitrariness or chance. Estimates of supervisors, as we have seen, vary from individ-

ual to individual and from trait to trait. Each factor which might contribute toward a correct statement based on criteria other than the tests themselves must itself be investigated, correlated with other criteria, and given its proportionate value in the trait or skill rating.

A Typical Investigation

The following series of results, from Rogers' study of typewriting and stenographic ability, serves to illustrate at the same time both the prognostic and the diagnostic use of this correlation method. A group of students, upon beginning their study of typewriting, were given various tests, chosen more or less at random. Each month thereafter ratings or measures of skill were obtained on the basis of set trials in the skill being acquired, with objective ratings for quantity and quality of work. On each monthly occasion these objective ratings were compared with the test standings, secured once for all at the beginning of the year. The table on the following page gives the correlation of test scores and current skill for each of the first three months and for the last month of the year's work:

CORRELATION OF TESTS WITH SKILL IN TYPEWRITING (ROGERS)

Test	October	November	December	April
Verb—Object ...	.41	.43	.46	.57
Color Naming....	.30	.43	.45	.61
Checking Numbers	.45	.47	.37	.30
Action—Agent ..	.42	.43	.29	.40
Substitution	.21	.27	.11	.42
Agent—Action ..	.29	.19	.40	.28
Direction	.11	.14	.19	.32
Opposites	.17	.11	.07	.54
Analogies	—.09	.21	.17	.00

It is clear, first, that these various tests correlated to different degrees with typewriting proficiency. They are arranged in the table in order of their significance throughout the year, as indices of this trait or capacity, from best to poorest. The correlation coefficients become smaller as one goes through the list of tests from Verb-Object to Analogies.

If we suppose that achievement after one school year of practice indicates with approximate accuracy the probable final standings in actual typewriting ability, it is clear that these significant tests are not only diagnostic at that time, but were also prognostic from the very beginning. Before the acquisition of trade skill was begun, these tests (the first four or five in the list as it stands) indicated with accuracy varying with their

respective correlations, the capacity of the individuals to acquire the trade skill in question. The remaining tests were less indicative or not at all significant of this capacity to learn. Although the correlations increase somewhat from month to month in all cases, it is mainly the tests that indicate relative aptitude in the first month that also indicate it most definitely at the end of training.

Initial scores in some of these tests seem to show no dependable relation to typewriting proficiency. They will not be included in the final team to be used for such prognosis or diagnosis. So far as the table gives information, it may well be that several of the tests give equally good correlations merely because they sample or reflect practically the same elements in the total pattern. Or they may alike depend on some general type of competence. The best team will be made up of those tests from this list or from lists derived from further investigation, all of which give significant but usually imperfect indication of the total pattern, but which, by not correlating closely among themselves, suggest that they represent different features of that pattern. Under refined

statistical treatment the precise amount of new information given by the addition of the results of any one test to those of another will be determined, and appropriate weight given to each test score, in terms of its independent contribution. The explanation of these statistical refinements is beyond our present purposes. It should not require pointing out that even these results, until further verified, apply only to typewriting ability under the circumstances described and measured by the criterion actually used on this occasion.

Differential Tests

That some specific pattern of traits or aptitudes is involved in typewriting is indicated by comparison of the correlation of these tests in the two cases of typewriting and stenography or grammar. In the following tabulation of Rogers' data the correlation of each test with typewriting at the mid-year period are compared with the correlations of each test with mid-year instructor's grades for stenography and for grammar, for the same group of students:

CORRELATION WITH MID-YEAR STATUS

Test	Typewriting	Stenography	Grammar
Verb—Object	.55	.36	.37
Color Naming	.41	.34	.38
Checking Numbers..	.53	.07	.22
Agent—Action	.31	.19	.37
Substitution	.37	.40	.16
Agent—Action	.31	.19	.37
Directions	.13	.46	.54
Opposites	.15	.45	.40
Analogies	.25	.31	.43

Here it appears that the same tests do not, in the main, correlate best with all three skill patterns. If we draw a line across the table separating the better from the poorer tests for typewriting, it is to be observed that although the good tests for this skill lie above the line, the tests that correlate well with the two other skills are found on the whole below the line. The typewriting correlations tend to decrease as one goes down the list of tests. Correlations with Stenography and with Grammar, on the other hand, tend to increase as one goes down the list of tests. It is just those tests that stand low as indices of typewriting skill that stand high as signs of stenographic and grammatical excellence. Certain tests, moreover, such as Color-Naming and Verb-Object, tend to correlate with all three skill patterns. These sug-

gest the presence in all these skills of a common factor.

A further illustration of the possibility of finding tests with specific differential value is to be found in an investigation reported by Link.[1] Production records of inspectors and gagers in a factory were correlated with five different tests. These tests were card sorting, tapping, cancelation, number group checking, and a directions test used as an index of general intelligence. The correlations were as follows:

DIFFERENTIAL CORRELATIONS (LINK)

	Sorting	Tapping	Cancelation	Checking	Intelligence
Inspectors	.55	.14	.63	.72	.14
Gagers	.05	.52	.17	—.10	.18

The single intelligence test does not correlate closely with production records in either operation. Of the remaining four tests, three correlate closely in the case of inspectors, the average of the three correlations being .63. But these three tests do not correlate with production in the case of the gaugers, the average of the three correlations being .01. The other test, Tapping, on the other hand, correlates fairly well in the case of gagers (.52), but poorly in the case of inspectors (.14).

[1] H. C. Link. *Employment Psychology*, p. 35.

Here, in other words, is one test that correlates with neither type of work, several that correlate with one or the other but not with both, and none that correlates significantly with both operations.

A still more striking case of the differentiating of trade skills by tests is reported by Bregman.[2] Sales girls and clerical workers were rated by supervisors so that they might be fairly classified as good, average, and poor, in both cases. In the case here considered the average workers are not considered, but test records of good and poor workers in each group are correlated with the ratings. The following table represents only a sample of the much more elaborate data presented in Bregman's paper, in which will be found also a description of the various tests employed.

CORRELATIONS WITH RATINGS

Test	Good and Poor Sales Girls	Good and Poor Clerical Workers
1	—.16	.09
2	—.66	.00
3	—.79	.31
4	—.22	.28
5	—.79	.37
6	—.54	.28
Total 1—6	—.59	.34

[2] Elsie Oschrin Bregman, "A Study in Industrial Psychology-Tests for Special Abilities," *Journal of Applied Psychology*, June, 1921.

The point of interest here is not so much the actual magnitude of the correlations as their contrasting direction. A set of tests which correlates positively with one operation correlates negatively with the other. The differential value of such tests is apparent. As Bregman remarks "what is apparent is the tendency for sales clerks and clericals to pull in distinctly opposite directions, so that we have a difference between trades as distinct as the differences within a trade, if not more so."

The examples given are typical of many now on record in the technical literature of mental measurement. References to several of these are given in the bibliography for readers who may be interested in more detailed acquaintance with this method of analysis. It is a method which, beyond all those we have considered, calls for statistical expertness, from the beginning of an empirical inquiry to the practical application of the results. Thus, so long as correlations of skill pattern or trait with test scores are not perfect, inference from test rating to most probable skill or trait rating, either prognostic or diagnostic, is not straightforward and simple. Since no perfect correlations have yet resulted from the use of this

method, the inference from test score to valid judgment of character calls for statistical technique no less rigorous than that involved in the selection of the significant tests.

It should of course be borne in mind that two traits, or a given trait and occupational success, may not correlate throughout their whole range. Thus intelligence below a certain point may disqualify one for the work of errand boy. Intelligence above a certain point may act in the same way, if only because of unrest and brief job tenure. But within the middle range, intelligence may be fairly well correlated with success in such work. A certain completeness of color vision is required to be a competent locomotive engineer, since colored light signals must be correctly identified. But beyond a certain point it is improbable that fineness of color discrimination is of any advantage in such work. Thus along a certain limited range of the degrees of a trait or of test scores, correlation may be close, although the total range may not disclose this fact, and single correlation coefficients for the whole series may convey false impressions.[3] For such reasons the

[3] E. L. Thorndike. "Fundamental Theorems in Judging Men," *Journal of Applied Psychology*, March, 1918.

correlation method may frequently be supplemented to advantage by other methods, such as those in which scatter diagrams of all the candidates are prepared, or by those in which psychographic analysis of the individual is made, or by those in which critical scores are determined. Some of these possibilities are indicated in the last chapter of this book.

The following classification will suggest the variety of uses to which the correlation method has been put in the search for methods of vocational and industrial analysis. Only those cases are cited in which correlation results in some form are given, and in the main only those for which published accounts are available. Reference to the definite sources will be found in the bibliography for this chapter in Appendix A.[4]

Clerical Workers. Thorndike, Dück, Link, Thurstone, Bregman, Hollingworth and Poffenberger, Burtt.

Typewriting. Lahy, Lough, Rogers, Link, Chapman.

Comptometers. Link, Marcus, Bills.

[4] A good summary account of many of these cases is given in Muscio, *Vocational Guidance* (A Review of the Literature), which appeared as Report No. 12, of the Industrial Fatigue Research Board, London, 1921.

Engineering. Thorndike, Thurstone, Scott (Trabue).

Music. Seashore.

Printing. Lipmann, Krais.

Salesmanship. Scott, Hollingworth, Oschrin, Muensterberg (Burtt).

Telegraphy. Jones, Thurstone.

Telephone Operating. Muensterberg, McComas, Nead, Fontegne and Solari.

Motormen. Muensterberg, Stern, Sachs.

Aviation. Head, Stamm, Henmon, Dockeray and Isaacs, Stratton-McComas-Hoover and Bagby.

Stenography. Lough, Rogers, Bills, Burtt.

Factory Operations. Link, Burr, Otis, Burtt, Carney, Allen, Hollingworth and Poffenberger.

As has been suggested already, the use of tests and of correlation methods in character analysis has distinct limitations, and in incompetent hands such methods may result in actual mischief. There are two ways, in practice, in which such dangers may best be avoided. If an industry or business or institution is sufficiently large to maintain its own personnel organization, a department of education, or a department of industrial relations,

at least one responsible member of this staff should be an expert trained in the methods of mental measurement, trade test procedure, correlation technic, and general statistical procedure. Where such specialization is not feasible, part-time consulting connections may usually be made with psychological and statistical specialists connected with university faculties, or with special bureaus of industrial research, such as the Psychological Corporation, with its various local branches, and the Bureau of Personnel Research of Carnegie Institute of Technology.

The empirical procedure is a method of character analysis easily open to misuse. Misuse of the method, through failure to comprehend its inevitable complexity, may easily result in professional chagrin and practical disappointment. Altogether, the method itself is more important than any particular results that have been achieved through its application. It is the final method that must be applied to test the validity of any presumed criteria of character. Especially from the point of view of character judgments in vocation and industry, the method has a twofold advantage. At the same time that it identifies the traits of

the individual that signify successful work, it indicates, although it may not name, the aptitudes which the execution of that work involves. It is thus at the same time a method of character judgment and a technic of job analysis.

CHAPTER XII

THE DIAGNOSIS OF TEMPERAMENT

Moral and Social Traits

That people differ in what we may call temperament, as well as in competence and in aptitude, in trade skill and in educational status, is apparent. If aptitude and interest determine what they do, and if competence sets a limit to their achievement, there is still to be considered their manner of doing. Two workmen of equal general competence, with identical degree of special skill, will nevertheless differ in character. One will work calmly, the other more excitedly; one will be steady, the other more erratic. Confidence and distrust, cheerfulness and gloom, generosity and selfishness, courage and cowardice, loyalty and infidelity, adaptability and stubbornness, truthfulness and deceitfulness, aggressiveness and submissiveness, taciturnity and loquacity, scepticism and credulity, and dozens of similar word pairs, indicate the extremes of vari-

ous lines drawn through human character in common speech.

In mental disease the constitutional attitudes, moods and dispositions often show themselves in heightened degree, and the volitional and emotional features may constitute essential aspects of the clinical pictures. The ancients were wont to attribute these constitutional differences to the balance of various bodily fluids in the individual's system, the bile, phlegm, blood, or to the mixture of elements such as earth, air, fire and water. Many modern authorities believe that some at least of these character traits depend upon the activity of various glands and their secretions. A recent writer suggests that temperament might significantly be called the "chemique" of the individual as distinguished from his "physique." Other authorities are inclined to attribute these moral or social qualities to the pattern of instinctive equipment inherited by the individual from his forebears. Some of the temperamental traits, again, seem to be explicable in terms of early established habits. Others may represent only special aspects of intelligence, or interests occasioned by special aptitudes. In a preceding section, moreover, it was suggested that many of

these terms refer not so much to traits of individuals as to the conduct of particular human pairs, or characteristics of human nature under special circumstances. Link, in his interesting book, "Employment Psychology," devotes considerable space to a suggestive discussion of the "relativity" of these moral and social characteristics of people. The following quotation indicates the general tenor of this author's point of view:

One of the great errors which employment managers, foremen, superintendents, and all other people, including ministers, teachers, and religious workers fall into, is the belief that the moral qualities are *absolute* qualities. They believe that if a man is lazy he *is* lazy. If he is industrious he *is* industrious. If he is cheerful he *is* cheerful. If he is disloyal he *is* disloyal. If he is ambitious he *is* ambitious. If he is good he *is* good. And if he is bad he *is* bad. In other words, they labor under the belief that the moral qualities are constant qualities which are an inseparable part of a human being as scales, fur, and hide are an inseparable feature of the fish, the dog, and the elephant; and further, that no matter where people are and what they are doing, their moral qualities are an invariable part of their nature. Nothing could be farther from the truth. The moral qualities are not absolute. They are not blanket qualities which cover an individual's entire range of life no matter under what circumstances he may live. On the contrary, moral traits are *relative,*

and their nature depends upon a very wide variety of external economic, social and bodily conditions.[1]

With all these qualifications and distinctions before us, we must necessarily speak of temperamental traits with considerable reserve. Trait-terms for these aspects of character are abundant, but knowledge concerning their elementariness, their independence, and their mutual relations is still to be acquired. Using these various terms, then, with a vagueness proportionate to our understanding, what progress has been made in the measurement of those characteristics to which they are intended to refer? What advance is to be reported beyond the loosely organized traditional methods described and criticized in earlier sections?

The concrete rating scales, utilizing personal specimens or graphic devices such as those advocated by Scott, Miner, the Committee on Classification of Personnel in the Army, and others, we have already had occasion to describe. These, however, do not constitute definite methods of diagnosis or measurement. They are, instead, methods of securing more systematic and consistent reports of the subjective impressions of

[1] Link. *Employment Psychology*, p. 203.

associates. They do not contribute information not already available in testimonial reports, although they may lend definition and uniformity to this evidence. Efforts to approach more closely to the technic of test procedure may best be presented by brief summaries of representative studies. Such surveys are given in the following paragraphs, followed by a general statement of the present status of prognostic and diagnostic methods in this field.

The Psychoneurotic Inventory

An interesting method has been devised by Woodworth in the effort to secure an index of the individual's "emotional stability," balance, or freedom from tendencies that predispose to nervous breakdown. An inventory of complaints or difficulties is provided, in the form of questions to be answered by "Yes" or "No." Each question relates to some tendency, habit, condition, experience, trait, or antecedent which, if answered in a particular manner, specified in the scoring card, indicates the individual to be "peculiar" in that respect. To be peculiar in this sense is to have the trait which most people lack or to lack that which most have. The immediate interest of the inven-

tory is not in the particular items reported, but rather in their number. The average or normal individual has of course a certain number of "peculiarities." College students and white army recruits, for example, were found to have an average of about ten such peculiarities, the largest possible score being over one hundred. Negro recruits were found to have larger scores than white recruits, twice as many in fact. Soldiers with nervous disorders were found to give still higher scores, the averages differing for various types of mental and nervous complaint. In one condition the average score of men in a certain hospital before the armistice was as high as forty points, and individuals with as many as seventy-five points were sometimes found.

This questionnaire method of indexing the degree of wholesomeness, the lack of "peculiarity" in nervous and mental health, we may call the Psychoneurotic Inventory. It is now being improved, adapted, and more fully tried out, in the hope that there may develop from it a useful means of expressing a very important temperamental characteristic. The inventory, in one of its early forms, has been reproduced with Professor Woodworth's permission, in the writer's

book, *The Psychology of Functional Neuroses.* Perhaps the greatest difficulty in the use of such an inventory, by others than the individual concerned, is the degree to which the replies vary with the motivation of the occasion. Scores made by neurotic soldiers after the armistice were strikingly lower than were those from similar patients before this event. The change in motivation from one toward invalidism to one toward recovery and discharge produced a more optimistic rendering, not only of present status but even of the facts of personal history. The method is essentially that of self-estimation, and the liability of self-estimates to distortion we have already had occasion to consider.

The Significance of Handwriting

Various psychologists have subjected handwriting to analysis and trial, in the endeavor to find in the characteristics of an individual's graphic performance some indication of more general trends. It is clear that the detailed rules of graphologists and of chirographic systems lead nowhere. They are dogmatically stated and without verifiable foundation. Hull and Montgomery submitted ten of the typical traditional grapholo-

gical criteria to test, by comparing these details of handwriting with combined judgments of associates on the traits professedly indicated. The correlations were only such as might result from chance arrangements. But such graphologists as have submitted to objective test, as in Binet's study, for example, seem to be more accurate than are their systems. Binet found that amateurs and professional graphologists alike showed more than chance success in indicating the relative intelligence of the writers of paired specimens. Even in judging their relative morality the successes were more frequent than chance would produce, and Binet was led to suggest that handwriting might yet be shown to afford valuable information in character diagnosis.

The most elaborate studies in this field are those of Downey, who finds various significant relations between certain features of the handwriting pattern and the more general pattern of the individual's mental and motor attitudes. On the basis of her experimental findings, Downey has formulated a series of twelve tests, designated the "Will Profile" method, most of the tests in which utilize handwriting. "The series includes speeded, re-

tarded, disguised, blocked, and automatic handwriting, slow and rapid imitation of script, and speeded writing in a restricted space. In many cases the reaction from this set of tests is somewhat definitely patterned. A relatively high score on the first four tests indicates a quick, flexible reaction; on the second four traits, it suggests an aggressive reaction; on the last four, a deliberate, methodical, careful reaction."

Through comparing test results, in an experimental manner, with known or estimated personal types, this author is led to believe that the "Will-Profile" "has considerable characterological significance and that it can be used to advantage not only in getting the general temperamental pattern of an individual but also in determining the specific combination of traits." In conjunction with intelligence tests "it certainly affords, in many situations, a basis for conservative prophecy." This "Will-Profile" method, based mainly on handwriting and its characteristic patterns, is now being tried out by other psychologists, and until it has stood the test of trial, criticism and modification, its value in character diagnosis cannot finally be judged.

Tests of Aggressiveness

Another investigator, on the trail of character traits, is Moore, who has tried out a variety of devices. One set of tests advocated by him is for the measurement of "aggressiveness." The series includes tests of eye control in personal interview; distraction, while adding, by staring, by electric shock, by a snake; and association responses to certain critical stimulus words. Use of the method with college students leads him to "believe that this test approximates a true measurement of aggressiveness more nearly than does the Army Alpha examination approximate the measurement of intelligence."

The unaggressive subjects, while adding a standard series of numbers mentally, were three times as likely as the aggressive ones to be considerably distracted by staring or by the presence of a snake, and more liable to distraction by electric shock. The very aggressive men were four times as likely to be positive and definite in their responses to "enterprise" and "success," twice as likely to give an energetic type of response to "company"; and only one-fourth as likely to re-

spond definitely and vividly to "danger" and "death," as the very unaggressive men.

The most important test, that of eye control, counts for one-half of the total possible score. In this test the candidate is "required to perform a somewhat difficult series of mental additions while constantly returning the fixed gaze" of the examiner. He is instructed that "under no circumstances should he let his gaze wander from that of the man facing him." While the subject works at addition, count is made of the number of times his gaze wanders and his eyes leave those of the examiner. College men, chosen by associates and faculty for estimated aggressiveness, show ten times as much eye control as do those similarly chosen for lack of this trait. The median score of the aggressive group is "no movements"; that of the unaggressive group is "five movements." The thirteen aggressive men made a total of only six movements; the thirteen unaggressive made a total of seventy-two movements.

"Thus the simple behavioristic fact of the ability to look another person in the eye seems to have such a high significance regarding the presence or absence of aggressiveness as to warrant giving it an extremely prominent place in any scoring

method devised as a measure of this trait. The correspondence is in fact so close as to justify the generalization that a stop watch and a pair of fixed eyes are the only indispensable laboratory equipment necessary for estimating roughly the degree of aggressiveness in at least four-fifths of the subjects."

Queries and objections will surely occur to the reader's mind in considering the use of such a set of tests, and these difficulties will emphasize the extraordinary obstacles encountered by any attempt to find valid measures of temperamental character traits. Thus the tendency to stare at a companion or at a speaker may be one of the partial facts that leads one's fellows to classify him as aggressive. Part of the correlation with the staring test would then be spurious. On the other hand, we have insisted that character is what it manifests itself to be, and if the shifty eye really marks a man as a weakling in the eyes of his fellows it is well to single this factor out of the total impression and to take diagnostic advantage of it. A man's reputation is not nearly so remote from his character as the copy books would have us believe.

The Association Methods

A type of association test, similar to that given as an example of the method of Response Values, in Chapter VII, has been used by several investigators, in the hope that it might significantly reveal the characteristic intellectual and emotional attitudes of the persons tested. In place of scoring the responses according to the frequency with which they are given by others, notice is taken of the nature of the relation between stimulus word and reaction word. On the basis of these relations several "types" of reactors have been described. Thus Jung insists that the characteristic association responses of different people in such a test do not really distinguish *intellectual* types, but rather *emotional* state. "Educated test persons usually show superficial and linguistically deep rooted associations, whereas the uneducated form more valuable associations and often of ingenious significance. This behavior would be paradoxical from an intellectual viewpoint. The meaningful associations of the uneducated are not really the product of intellectual thinking, but are simply the results of a special emotional state. The whole thing is more important to the uneducated, his

emotion is greater and for that reason he pays more attention to the experiment than the educated person, and that is why his associations are more significant."

Jung then points out several types or sub-types. There is said to be an *objective* type, with undisturbed and usual reactions, and a *complex* type with many disturbances occasioned by special emotional complexes or sore points. There is a *definition* type described which "consists in the fact that the reaction always gives an explanation or a definition of the content of the stimulus word." Persons belonging to this type are said to wish to be more than they are, and such reaction tendencies are said to characterize stupid persons but also persons who are not stupid but have an "intelligence complex" and are fearful of being thought stupid. In another suggested type, the *predicate* type, it is not the intellectual but the emotional significance of the stimulus word that is prominent in the response, and reaction words indicating strong personal evaluation occur. "Just as the definition type really conceals a lack of intelligence, so the excessive emotional expression conceals or overcompensates an emotional deficiency," says Jung.

The "predicate" type of response in such free association tests has also been called the "egocentric" type. "The number of these 'egocentric' associations," says Wells,"[2] "has been thought, with reason, to bear a peculiar relation to the subject's general personality." "In normal persons, from 15 to 45 per cent of associative responses belong to this group. Single series have been taken with as low as 2 per cent and as high as 60 per cent; but the number of these is a fairly constant attribute of the individual." Excessive egocentric responses are believed to indicate either a specific or a general maladjustment in the individual's affective or instinctive life, and therefore to signify a lack of temperamental balance. Wells has reported interesting experiments designed to test out this belief, and concludes that his general results "are consistent with the view that the egocentric association type indicates a greater 'loading' of the experiment with affect; and, as the affect is there to be loaded, it is evidently not taken care of elsewhere in the personality."

In similar ways the association test has been used by many investigators as giving a possible

[2] F. L. Wells. *Mental Adjustments*, p. 261 ff.

clue to temperamental traits—cheerfulness or gloominess as indicated by the predominance of pleasant or unpleasant associations; objective attitude or introversion as indicated by the relative numbers of common and individual reactions; the relative strength of instincts and interests, as indicated by the speed of reaction to certain words, the vividness of certain responses, and similar criteria. However suggestive these results may be in striking cases of mental disorder, and however truly they may reveal the special mental set or attitude taken by the subject on the occasion of the experiment, the advocates of the verbal association technic are still far from providing measures of temperament. But many psychologists still look for definite advance in the application of this method. In recent work for example, by specialists in the use of the method in various forms, it is held to be "in the foremost rank among the methods of individual psychology," and, in one of its forms to give "perhaps the best objective correlate of temperament at present to hand." But this after all, in the light of our consideration of the validity of other methods of temperamental diagnosis, cannot be considered immoderate praise.

Measurements of Suggestibility

That people differ in the readiness with which their conclusions are influenced by the opinions or suggestions of others, is a matter of common observation. For this characteristic the term "suggestibility" is often employed. Stern has defined this tendency as "the imitative assumption of a mental attitude under the illusion of assuming it spontaneously." What is meant by suggestibility, in this sense then, is not willingness to be convinced by adequate evidence, openness to argument and reason, but what Whipple has called "uncritical acceptance of a notion, usually with the implication that the suggested individual is unaware that his ideas have been thus affected." Many tests involving suggestibility of this kind have been devised, in which through some verbal, gestural or circumstantial suggestion, the individual is led to react otherwise than he would if left to himself. Several investigations have employed such tests, in the endeavor to determine whether or not degree of suggestibility is a general characteristic; whether, that is, it is present in something like equal degree in all of an individual's conduct. The outcome of these investigations is with great

uniformity of a negative sort. Individuals differ in any given suggestibility test, but their differences are not carried over to other tests as fixed personal differences. In other tests they differ again, but in new ways and to quite new degrees. The tendencies underlying suggestibility seem then to be specific to the situation in hand; they characterize the individual under special circumstances rather than the individual as a person. Diligent search and experiment have not succeeded therefore in discovering any tests that can be said to measure the general suggestibility of individuals, for such a trait does not seem to exist.[3]

Detecting Moods and Emotions

Washburn and her students, in seeking for indications of moods or temperaments of cheerfulness and depression, employed the following instructions in a modified association test: "When I pronounce a word to you, observe what idea that word first calls to your mind and report whether it is a pleasant or an unpleasant idea. If it seems neither pleasant nor unpleasant, but indifferent,

[3] Interesting descriptions of typical tests involving suggestion are to be found in Whipple's *Manual of Mental and Physical Tests*, and in Brown's *Individual and Sex Differences in Suggestibility*.

continue thinking until either a pleasant or unpleasant idea is suggested, and report what it is." Having determined, under the conditions of the experiment, the average tendency to report pleasant ideas for the series of stimulus words, deviations from this average were used to indicate tendencies toward cheerfulness or depression. Comparison of such results with estimates by intimate associates gave definite positive correlations. This promising lead is reported as now undergoing the further investigation which it requires before its general diagnostic value is known.

A series of five tests intended to indicate tendencies toward emotional reactions and idiosyncrasies in this respect has been employed by Pressey. The subject is required first to cross out the words in lists provided, which name unpleasant things, and to indicate in each line which word is most unpleasant. The number of unpleasant words and the number of most unpleasant words not so marked by most people, are taken to indicate tendency to emotionalize and to be emotionally peculiar. Crossing out irrelevant words inserted in dull and exciting passages of reading matter, is used to give a measure of the tendency to absorption in the more exciting pass-

age. In lists of acts of varying degrees of social sanction, the "worst" acts and the "most common" acts are to be indicated. Deviation from the common responses is used as a measure of moral discrimination and experience. A modified form of the free association test gives a measure of the tendency to select common associates. A test for memory of emotional and unemotional words used in the preceding tests gives, finally, a measure of "emotional memory." This test is still in very tentative form, and its value therefore undetermined.

General Conclusions

A variety of other devices might be cited as having been, in one form or another, used in the attempt to measure or analyze or identify temperamental traits. Self-analysis charts, sometimes with roughly quantitative scores suggested; graphic schemes for the visual portrayal of relative prominence of different traits as thus estimated; questionnaires, directing attention and comment to what are supposed to be significant indications of interests and attitudes; lists of interests and activities, for checking or for serial arrangement; analysis of dreams, with the intent

to discover the trends and motivations of the individual's thought; estimates, by acquaintances, of the individual's possession of specified characteristics; rating scales for more systematic expression of such personal estimates; dogmatic inference from physical structure or contour to temperamental quality; and finally, the various traditional methods which we have considered in preceding chapters; these indicate the variety of technics often utilized for temperamental diagnosis. The instances we have surveyed constitute some of the more definitely formulated and experimentally conceived endeavors.

On the whole then it is clear from the survey given that the measurement of temperamental traits is still in a preliminary experimental stage. Many aspects of human character need no longer be merely judged, no longer merely guessed at by traditional technics. General competence or intelligence, mental maturity, several special aptitudes, numerous trade skills, educational status and accomplishment, various types of vocational fitness, range of information in various fields, can be objectively measured either in all or in some of their features. Until the diagnosis of temperament has passed through its earlier ex-

perimental stages, the moral and social qualities, the interests, emotional dispositions, the attitudes and the volitional and instinctive characteristics, must continue to be estimated in traditional ways. But, as we have seen, even the best of these methods admits of improved technic and more systematic application. The contribution of science toward the art of judging human character cannot be limited to the devising of tests and measures. Until that distant day when objective methods of diagnosis may be available for all the interesting features of the complex human personality, science is under the obligation of submitting even the roughest traditional methods to constructive criticism. Even when cripples cannot be cured, it is the task of science to improve the crutch.

CHAPTER XIII

PSYCHOGRAPHIC METHODS

In the literature of vocational psychology special consideration has been given to the "individual psychograph" and to the "vocational psychograph." By the former was meant the attempt to give a full account of the mental characteristics of a particular person, as in the studies, by Toulouse, of such distinguished men as Zola and Poincaré. By the latter term was meant what is now more commonly called "job analysis." In more recent developments the term "psychograph" has been used especially to designate a particular analytic and graphic method of exhibiting the measures of an individual in some array of tests or traits. The word "profile" has also been used to indicate the graph resulting from such a charting of measures. Some consideration of the nature, advantages and applications of the method of the profile or psychograph may usefully constitute the subject of our closing chapter. Reference will also be made to various actual cases in

which the psychographic method has been employed.

Psychographic Technic

This method requires that, in a variety of tests or traits, measures be available which are susceptible of expression in some common fashion or in identical terms. Any of the standard "expression units" may be used, with the usual exception of Original Scores. Thus the measures of all traits may be expressed in Distribution Units, or in Ranks or Percentile Units, or in Absolute Units, or in Developmental Units such as mental age, or skill level. Such measures may then, in fairly legitimate ways, be compared, combined, averaged, and otherwise treated in quantitative fashion. When such measures are available a simple scheme of graphic portrayal serves many useful purposes.

Vertical lines may be erected, for example, along a horizontal base line, at equal distances. Each such line or column may stand for one of the tests or traits. Along one of the verticals may be laid off a scale in terms of the units to be adopted. On each of the remaining verticals may then be indicated, by cross bar or other indication, the degree or amount of the particular trait, or the individ-

ual's standing in it. This array of points, sometimes connected by lines, gives a readily apprehended picture, not only of the individual's general level of performance, but also of his relative status in the various respects, and of the uniformity or balance of his equipment in the features measured.

A horizontal line drawn straight across the chart may indicate the average level of all the characteristics, a summary of the general status. Another such line, similarly drawn, may be used to indicate the expected or normal status. The proximity of these lines will thus indicate the individual's general approach to normal or expected status. Their relative positions will indicate either retarded or backward condition, on the one hand, or precocity and superior status on the other. A third line may, for special purposes, be used to indicate the level of some critical deviation, a deviation, for example, of sufficient degree to indicate practical inadequacy, rejection, promotion, institutional care, lack of testamentary capacity, and so on.

If the tests or traits are grouped in the chart, on the basis of their common resemblances or their relation to certain more general aspects of personality, subordinate lines may be drawn, indi-

cating status in these special sections or divisions. In case it is possible to arrange the traits or tests in a hierarchy on one or other basis, the general direction or slope of the profile line also becomes significant.

The degree of scattering of the separate scores or measures about their average line also has special significance in such a psychograph. It indicates the symmetry of the individual's characteristics. The theoretical "average person" would be at the expected or normal line in all the features represented. Instability of organization, unevenness of development, or marked special aptitudes and disabilities will be indicated by appropriate deviations of actual from expected or average status. Marked irregularity or scattering of scores is, it seems, likely to be associated in many cases with temperamental peculiarities. It may also, in special cases, indicate disease or deterioration, or the influence of particular factors such as drugs or such as highly specialized practice.

Typical Psychographic Profiles

In the following charts are shown, by way of

examples, the partial psychographs of three boys. All three boys have the same general or average competence. In terms of developmental age units, this is average fifteen-year-old capacity. Rated on the miscellaneous intelligence scales, all three will be found to have the same mental age.[1] But the psychographs show the abilities of the three boys to be quite differently patterned. One of the profiles starts high and gradually falls. Another starts low and gradually rises. The third maintains a fairly uniform horizontal course.

The tests employed in these three cases are arranged in a rough order, beginning with those that involve mainly motor behavior and the handling of concrete objects, actually present, through an intermediate region, to tests involving little overt muscular dexterity, but calling essentially for the mental manipulation of symbols, relations, meanings and abstract situations. It is clear that, so far at least as present status is concerned, these boys are very differently endowed. The one excels his own average in the more mechanical

[1] Percentile or distribution units are in general preferable for psychographic purposes, since they are not so liable to misinterpretation. But since for some of the tests used on these boys, the only norms available are on a mental age basis, that method is here employed.

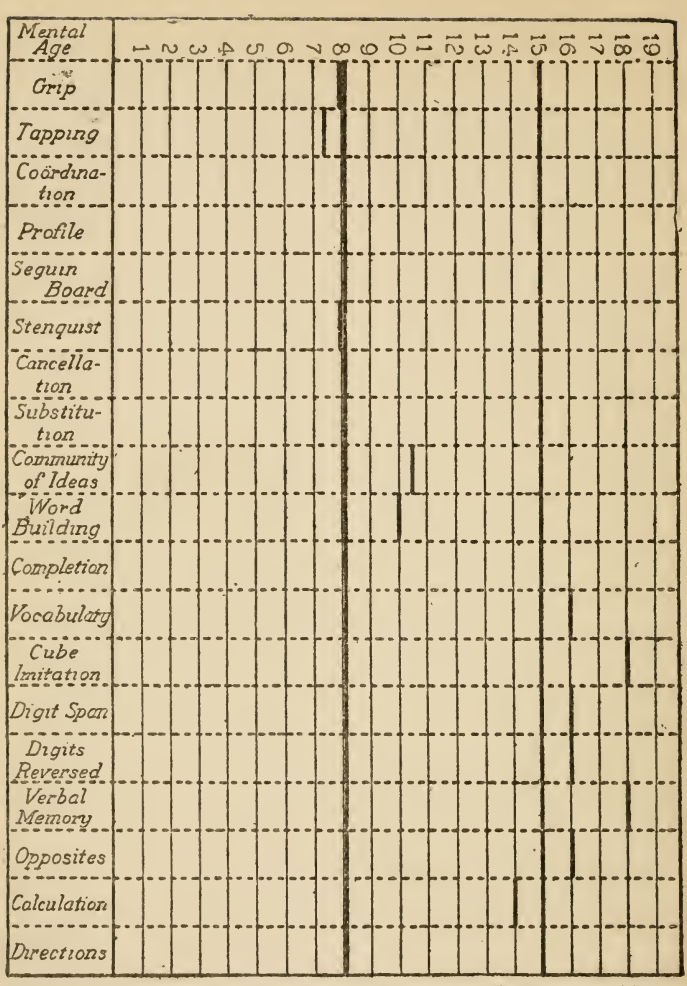


FIG. 2.—BOY A. Actual age, eight years. Mental age, fifteen years.

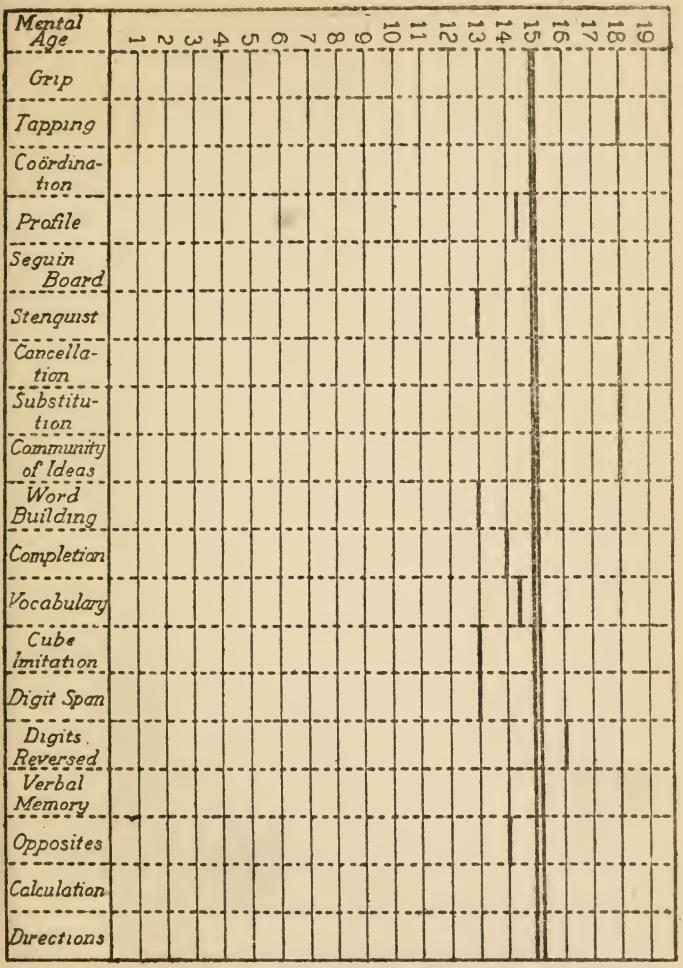


FIG. 3.—BOY B. Actual age, fifteen years. Mental age, fifteen years.

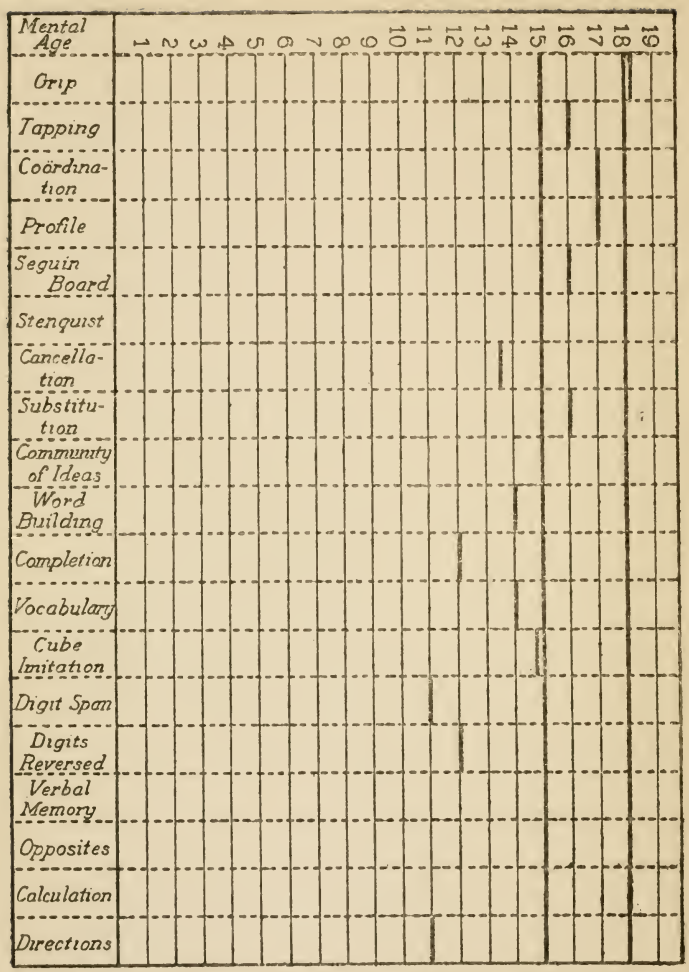


FIG. 4.—BOY C. Actual age, eighteen years. Mental age, fifteen years.

and manipulative skills; another excels his own average in more abstract and verbal operations; the third has no outstanding aptitudes nor disabilities, but is near his average in all the tests.

For purposes of immediate selection and placement, the facts up to this point are significant and applicable. From the point of view, however, of prediction and guidance or advice, certain further facts are of prime importance. Although these three boys have equal general competence at the date of examination, it is quite certain that this will not always be the case with them. Note that boy A is only eight years old, whereas B is 15, or nearly twice as old as A, and C, being 18, is more than twice as old as A. Our general knowledge of mental development, verified by experimental data, justifies the conclusion that A is only about half way through his chronological period of growth; that B has just about completed his growth in native capacity, and that C has already ceased to develop in this respect. The profile of C therefore indicates his permanent level of achievement; B may change somewhat but only slightly; but A will surely grow very materially in all respects, and perhaps unequally in the vari-

ous skills here represented. Boy A, in other words, is precocious; B is quite an average boy; C is dull and backward. These facts are shown by the relative positions and distances of the lines of average competence and of expected achievement. These samples serve briefly to illustrate the very useful type of analysis that is facilitated by the psychographic method. The chart is of course not absolutely necessary, since the figures tell the same story; but the graph presents the facts with a concreteness and vividness that the figures lack to all except those adept in their use.

Uses of the Profile Method

In recent years the profile or graphic method has often been adopted. Stern, many years ago, described the possibilities of some such technic. Rossolimo later used the graphic device, but for presenting data which were not expressed in adequately conceived units. Various professional "characterologists" have employed such charts for recording personal opinions and self-estimates in pseudo-exact terms. Allport has used a similar method for portraying personality profiles on the basis of questionnaire replies, with rough steps of classification. Downey, whose "Will-Profile"

method we have considered in another chapter, uses the graphic device for portraying and analyzing the pattern of the various handwriting and other reactions. Seashore, who has established a series of quantitative measures for certain aptitudes involved in musical ability, uses the graphic profile to show the general level and symmetry of these measures in the case of individuals who have been examined. Educational psychologists, as Courtis for example, employ a somewhat similar device for representing the balance of a pupil's achievements in various school subjects. Kitson has used the psychograph as a concrete method of exhibiting the standing of college students in various skills and traits. Yerkes and Cobb, treating separately the various tests which comprise the Alpha intelligence scale used in the army, have shown that medical men in the military service show a characteristic profile in these tests. The method, in its various forms, has distinct advantages as an instrument of mental analysis in private consultation, clinical examination, psychopathological research, in the re-education of inadequate personalities, and in experiments in vocational guidance and placement.

The full utility of the psychographic method will be accomplished only when technic and norms are available for the real measurement of a much larger array of characteristics. Such a personality picture has been for many years the goal of individual psychology. Kraepelin, writing nearly thirty years ago, expressed the need of such a method of portraying the "psychic status praesens." In psychopathological work, with which Kraepelin is chiefly identified, the need for more adequate exploration of the personality has always been real, and makeshifts have been frequently lamented. Kraepelin made many fertile suggestions toward the development of character measurements. Thus he commented on the Bertillon system of identifying an individual by his physical measurements and desired to develop a similar method of picturing an individual's mind. He insisted on the necessity of exact measurements, on the necessity of securing norms and measures of deviation, and he attempted to indicate some of the important "personal dispositions" or aspects of the personality. In the work of Kraepelin and his students are to be found the beginnings of many of the practical applications of experimental technic, graphic record, and

statistical procedure in character study. Even the scales for the measurement of educational accomplishment, so popular in recent years, are anticipated in Sommer's impressionistically arranged scales for determining ability in the fundamental processes of arithmetic.

This presentation of the psychographic methods concludes the general program to which this book is committed. We have reviewed the various traditional methods, indicated in some detail their various inadequacies, and suggested possible methods of increasing their value or at least of estimating their validity. The greatest hope seemed to lie in the supplementary or substitutive contribution of more accurate data through the more objective means of mental measurement. A survey of some of the principal methods of mental measurement disclosed a variety of technics and problems so complex as to be sufficient to constitute the subject matter of a definite professional activity. Readers who may be interested in a more detailed study of this field are provided, in the Appendix, with a representative but by no means complete bibliography, classified under the various chapter headings of the book. The signal importance of the problem of judging human character

is to be seen in the way in which all manner of interests have now begun to coöperate in its solution. The prospect of future progress in this field will be much increased by the continual active cooperation of such diverse interests as those represented by teachers, administrators, business men, psychopathologists, statisticians, managers, psychologists and personnel experts. The organized search, in our own generation, for fuller knowledge of human nature, may result in as great a contribution to human welfare as that which the past generation made to our knowledge of the physical world.

APPENDIX A

BIBLIOGRAPHY

I. General Introduction

Hollingworth, H. L. *Vocational Psychology* (Appleton).

Hollingworth and Poffenberger. *Applied Psychology* (Appleton).

Link, Henry C. *Employment Psychology* (Macmillan).

Scott and Hayes. *Science and Common Sense in Dealing with Men* (Ronald).

II. Letters of Application

Andrews, L. G. "A Grading System for Picking Men," *Sales Management,* January, 1922.

Personnel System of the U. S. Army, Vol. II, p. 604-633 (Superintendent of Documents, Government Printing Office, Washington, D. C.).

Poffenberger and Vartanian. "The Letter of Application in Vocational Selection," *Jour. Appl. Psychol.,* March, 1922.

Walton, Lillian C. "Study of Judgments of Letters of Application" (*Masters' Essay,* Dept. of Psychology, Columbia University).

Wells, F. L. "Analysis of a Successful Agent," *Life Association News,* Vol. XI, No. 3.

III. Human Character in Photographs

Anderson, L. D. "Estimating Intelligence by Means of Printed Photographs," *Jour. Appl. Psychol.*, June, 1921.

Hollingworth, H. L. *Vocational Psychology.* Chapter II.

Pintner, R. "Intelligence as Estimated from Photographs," *Psychol. Rev.*, July, 1918.

IV. Judging One's Own Characteristics

Allport, F. H., and G. W. "Personality Traits, their Classification and Measurement," *Jour. Appl. Psychol.*, April, 1921.

Cattell, J. McK. "Homo Scientificus Americanus," *Science*, XVII.

Hollingworth, H. L. *Vocational Psychology*, Chapters VI, VII.

Knight and Franzen. "Pitfalls in Rating Schemes," *Jour. Educ. Psychol.*, 1922.

Yerkes and LaRue. *Outline for the Study of the Self* (Harvard University Press).

V. The Personal Interview

Binet, A. "Nouvelles recherches sur la mesure du niveau intellectuel, etc.," *L'Année Psychol.* XVII, 1911, 9:182.

Link, H. C. *Employment Psychology.*

Scott, W. D. "Scientific Selection of Salesmen," *Advertising and Selling Magazine*, October, 1915.

VI. Recommendations and Testimonials

Folsom, J. K. "Statistical Study of Character," *Ped. Sem.*, September, 1917.

Hollingworth, H. L. *Vocational Psychology,* Chapters VI, VII.

Kohs and Irle. "Prophesying Army Promotion," *Jour. Appl. Psychol.*, March, 1920.

Norsworthy, N. "On the Validity of Judgments of Character," *Essays in Honor of William James.*

Simpson, B. R., "Reliability of Estimates of General Intelligence," *Jour. Educ. Psychol.*, April, 1915.

Slawson, J. "Reliability of Judgment of Personal Traits," *M.A. Essay,* Dept. of Psychology, Columbia University, 1920.

The Personnel System of the U. S. Army, 2 vols. (Govt. Printing Office, Washington, D. C., Section on Rating Scales).

VII. Common Sources of Error

Achilles, P. S. and E. M. "Estimates of the Military Value of Certain Character Qualities," *Jour. Appl. Psychol.*, December, 1917.

Hollingworth, H. L. "Experimental Studies in Judgment," *Arch. of Psychol.*, Columbia University, No. 29.

Miner, J. B. "Evaluation of a Method for Finely Graduated Estimates of Abilities," *Jour. Appl. Psychol.*, March, 1917.

Rugg, H. O. "Is the Rating of Human Character Practicable?" *Jour. Educ. Psychol.*, November, 1921, February, 1922.

TERMAN, L. M. *The Measurement of Intelligence* (Houghton Mifflin).

THORNDIKE, E. L. "A Constant Error in Psychological Ratings," *Jour. Appl. Psychol.*, March, 1920.

VIII. MEASURING MENTAL COMPETENCE

CHAPMAN, J. C. *Trade Tests* (Holt).

MITCHELL and RUGER. *Psychological Tests, Bibliography* (Bull. IX, Bureau of Educ. Exper., New York).

MONROE, DEVOSS and KELLY. *Educational Tests and Measurements* (Houghton Mifflin).

PINTNER, R. *The Mental Survey* (Appleton).

PINTNER and PATTERSON. *A Scale of Performance Tests* (Appleton).

PYLE, W. H. *The Examination of School Children* (Macmillan).

STERN, W. *Psychological Methods of Testing Intelligence* (Warwick and York).

TERMAN. *The Measurement of Intelligence* (Houghton Mifflin).

THORNDIKE, E. L. "Educational Diagnosis," *Science*, January, 1913. *Mental and Social Measurements* (Teachers College).

TRABUE and STOCKBRIDGE. *Measure Your Mind* (Doubleday Page).

WHIPPLE, G. M. *Manual of Mental and Physical Tests* (Warwick and York).

YOAKUM and YERKES. *Army Mental Tests* (Holt).

IX. GENERAL COMPETENCE AND SPECIAL APTITUDE

BRONNER, A. C. *Special Abilities and Disabilities* (Little, Brown).

HOLLINGWORTH, H. L. *Vocational Psychology,* Chapter IX.

LINK, H. C. *Employment Psychology,* Chapter XII.

MCCALL, W. A. *How to Measure in Education* (Macmillan).

SEASHORE, C. E. *Psychology of Musical Talent* (Silver, Burdett).

THORNDIKE, E. L. *Educational Psychology,* 3 vols. (Teachers College).

X. The Method of the Sample

CHAPMAN, J. C. *Trade Tests* (Holt).

HOKE and WILSON. *How to Measure* (Macmillan).

HOLLINGWORTH, H. L. *Vocational Psychology,* Chapter V.

MURDOCH, K. "Measurement of Certain Elements of Hand Sewing," *Teachers College Publications.*

STARCH, D. *Educational Psychology.*

THORNDIKE, E. L. "Handwriting," *Teachers College Record,* March, 1910. "Measurement of Achievement in Drawing," *Teachers College Record,* November, 1913.

XI. The Correlation Methods

ALLEN, G. L. "A Correlation Between Mental Tests and Manual Aptitudes," *Masters' Essay,* Columbia University, 1916.

BREGMAN, E. O. "A Study in Industrial Psychology-Tests for Special Abilities," *Jour. Appl. Psychol.,* June, 1921.

BURTT, H. E. "Prof. Muensterberg's Vocational Tests," *Jour. Appl. Psychol.*, September, 1917. "Employment Psychology in the Rubber Industry," *Jour. Appl. Psychol.*, March, 1920.

CARNEY, C. S. "Some Experiments with Mental Tests as an Aid in Selection and Placement," *Bull. Exten. Div.*, Indiana University, September, 1919.

CAROTHERS, F. E. "Psychological Examinations of College Students, *Archives of Psychol.*, Columbia University.

DOCKERAY and ISAACS. "Psychological Research in Aviation," *Jour. Compar. Psychol.*, April, 1921.

FLANDERS, J. K. "Mental Tests of a Group of Employed Men," *Jour. Appl. Psychol.*, September, 1918.

HENMON, V. A. C. "Air Service Tests of Aptitude for Flying," *Jour. Appl. Psychol.*, June, 1919.

HOLLINGWORTH, H. L. *Vocational Psychology*, Chapter V.

HOLLINGWORTH and POFFENBERGER. *Applied Psychology*, Chapter XI.

JONES, E. S. "The Woolley Test Series Applied to the Detection of Ability in Telegraphy," *Jour. Educ. Psychol.*, 8, 1917.

KITSON, H. D. "The Scientific Study of the College Student," *Psychol. Mon.*, 1917, 23, No. 1.

LINK, H. C. *Employment Psychology*, Part I

MARCUS, L. "Vocational Selection for Specialized Tasks," *Jour. Appl. Psychol.*

MUSCIO, B. "Vocational Guidance," *A Review of the Literature.* Report No. 12, Industrial Fatigue Research Board, London, 1921.

Oschrin, E. "Vocational Tests for Retail Saleswomen." *Jour. Appl. Psychol.,* June, 1918.

Otis, A. S. "Selection of Mill Workers by Mental Tests," *Jour. Appl. Psychol.*

Rugg, H. O. *Statistical Methods in Education.*

Rogers, H. W. "Psychological Tests for Stenographers and Typewriters," *Jour. Appl. Psychol.*

Scott, W. D. "Scientific Selection of Salesmen," *Advertising and Selling Magazine,* October, 1915. "Selection of Employees by Means of Quantitative Determinations, *Annals," Amer. Acad. Pol. and Soc. Science,* Vol. 65, May, 1916.

Stratton, McComas, Coover and Bagby. "Psychological Tests for Selecting Aviators," *Jour. Exper. Psychol.,* 1920, 3.

Terman, L. M. "A Trial of Mental and Pedagogical Tests in a Civil Service Examination for Policemen and Firemen," *Jour. Appl. Psychol.,* March, 1917.

Thorndike, E. L. "Tests for Vocational Selection," *Bull. Exten. Div.,* Indiana University. Vol. IV, 4. *Mental and Social Measurements* (Teachers College). "Fundamental Theorems in Judging Men," *Jour. Appl. Psychol.,* March, 1918.

Thurstone, L. L. "Mental Tests for Prospective Telegraphers," *Jour. Appl. Psychol.,* June, 1919. "A Standard Test for Office Clerks," *Jour. Appl. Psychol.,* September, 1919.

XII. Diagnosis of Temperament

Allport, F. H. and G. W. "Personality Traits, their Classification and Measurement," *Jour. Abn. and Soc. Psychol.,* April, 1921.

BROWN, W. "Individual and Sex Differences in Suggestibility," *Univ. of Calif. Pub. in Psychol.*, Vol. 2, No. 6, July, 1916

DOWNEY, J. E. "The Will-Profile, A Tentative Scale for Measurement of the Volitional Pattern," *Univ. of Wyoming Bull.*, Vol. 16, No. 4, 1912.

HOLLINGWORTH, H. L. *Psychology of Functional Neuroses*, 1920, Chapter VIII. *Vocational Psychology*, Chapter IX.

HULL and MONTGOMERY. "An Experimental Investigation of Certain Alleged Relations Between Character and Handwriting," *Psych. Rev.*, January, 1919.

JUNG, C. G. "The Association Method," *Am. Jour. of Psychol.*

KENT and ROSANOFF. "A Study of Association in Insanity," *Amer. Jour. of Insanity*, Nos. 1 and 2, 1910.

LINK, H. C. *Employment Psychology*, Chapters XVI, XVII.

MCDOUGALL, WM. *Social Psychology.*

MOORE, H. C. "The Measurement of Aggressiveness," *Jour. Appl. Psychol.*, June, 1921.

PRESSEY, S. L. and CHAMBERS, O. R. "First Revision of a Group Scale Designed for Investigating the Emotions," *Jour. Appl. Psychol.*, March, 1920.

WASHBURN, M. F. and others. "An Attempt to Test Moods or Temperaments of Cheerfulness and Depression, etc.," *Amer. Jour. Psychol.*, July, 1919.

WELLS, F. L. *Mental Adjustments.* Chapter VII.

WHIPPLE, G. M. *Manual of Mental and Physical Tests*, Vol. II, Chapter X.

WOODWORTH and WELLS. "Association Tests," *Psychol. Monog.*, XIII, 5, 1911.

XIII. The Psychographic Methods

Allport, F. H. and G. W. "Personality Traits, Their Classification and Measurement," *Jour. of Abn. and Soc. Psychol.,* April, 1921.

Courtis, S. A. "Report of Second Conference on Educational Measurements, Indiana University," p. 153, *Indiana Univ. Bull. XIII,* 11, October, 1915.

Downey, J. E. "The Will Profile," *Univ. of Wyoming Bull.* XVI, 4, 1912.

Hollingworth, H. L. *Vocational Psychology,* Chapter IV. *Psychology of Functional Neuroses,* Chapters VIII, XIV.

Kitson, H. D. "The Scientific Study of the College Student," *Psych. Mon.,* 23, No. 1, 1917.

Kraepelin, E. "Der Psychologische Versuch in der Psychiatrie," *Psychologische Arbeiten,* 1, 1894.

Parker, B. "The Psychograph of Rossolimo," *Amer. Jour. Insan.,* 73, 1916.

Rossolimo, G. "Die Psychologische Profile," *Klinik f. psychische u. nervöse Krankheiten,* 6, 1911, Heft, 3, and later numbers.

Seashore, C. E. *Psychology of Musical Talent.*

Stern, W. *Die Differentielle Psychologie,* 1911.

Stevens, H. C. "A Revision of the Rossolimo Tests," *Titchener Commem. Vol.,* 1917.

Whipple, G. M. *Manual of Mental and Physical Tests.*

Yerkes and Cobb. "Intellectual and Educational Status of the Medical Profession, etc.," *Bull. of Nat. Res. Council,* 1, 8, No. 8, February, 1921.

Books on Statistical Methods

Brinton, W. C. *Graphic Methods for Presenting Facts.*
McCall, W. A. *How to Measure in Education.*
Rugg, H. O. *Statistical Methods in Education.*
Secrist. *An Introduction to Statistical Methods.*
Thorndike, E. L. *Mental and Social Measurements.*
Whipple, G. M. *Manual of Mental and Physical Tests.*
Yule, G. U. *Introduction to the Theory of Statistics.*

APPENDIX B

LABORATORY EXERCISES

Experiments and Laboratory Exercises are here given which will illustrate and clarify the topics and methods discussed in the foregoing chapters. The author finds it instructive and convenient to use these exercises or adaptations of them, in college and university courses in Mental Measurement, or in Vocational and Industrial Psychology, or in Applied Psychology. For such courses the present book serves as a convenient elementary text, and may be supplemented to any desirable degree by more detailed lectures and demonstrations, and by readings or reports of books and articles suggested in the preceding Bibliography.

EXERCISE I

Letters of Application

A series of letters of application for some position may readily be secured, in various ways. Thus a business man who has advertised for an assistant, a newspaper that has run classified want advertisements, a family that has advertised for a housekeeper or servant, a school or college that has received applications for admission, may be willing to provide such a series for scientific use. All such material should of course be kept strictly confidential, and used only for experimental purposes, unless the writers' consent to the use of their letters is secured.

With the letters of application available, choose a representative set of fifteen letters. Keep each letter separate and intact, enclosed in its envelope, with the stationery and original penmanship carefully preserved. Place on each a key letter, as A, B, C, D, etc., which will identify the letters and envelopes.

Considering the position for which the writers have applied, select four important traits, which should be possessed by satisfactory applicants for this position. They may, for example, be such traits as Honesty, Intelligence, Penmanship, Tact, Leadership, Ambition, Neatness, etc.

Ask ten different individuals to arrange the set of fifteen letters in an order of merit for one or more of



A



B



C



D

these traits, placing in first place the best, in second place the next best, etc.

After two weeks or so have elapsed, ask the same individuals again to arrange these letters for the same trait or traits.

Now prepare a table of results similar to those on pages 13-14 showing the amount of agreement between different judges in their first arrangement. Find the average position assigned each letter, and the deviations of the ten judges from this average. This will show, when these deviations are averaged, how much the judges tend to disagree on each letter. Averaging all of the average deviations will show how much the judges tend to disagree on the trait in question.

Do the judges agree more closely on some letters than on others? What are the characteristics of such letters? Do the judges agree on some traits more closely than on others? What are these traits?

Compare, in the case of each judge, his arrangements of the letters for a given trait on two occasions. In what simple way can you get a measure of his consistency, in each case? Are some judges more consistent than others? Are judges in general more consistent on some traits than on others? What explanations have you to offer for the results? How do your results compare with the conclusions presented in Chapter II?

EXERCISE II

Human Character in Photographs

Photographs of twelve children are here provided. Each photograph is indicated by a letter. Arrange the letters in a column alphabetically, so that after each letter may be placed numbers indicating the position assigned the corresponding photograph by each of several judges.

Select certain interesting character traits for judgment, such, for example, as Humor, Honesty, Aggressiveness, Intelligence, etc. If Humor, for example, is selected, ask various persons to indicate the order of merit of the individuals whose photographs are shown, ranking them from 1 to 12, Number 1 being the highest in the trait, Number 2 the next, and so on. In this way judgments may be made for several traits.

When several judges have made their rankings, make a tabular report of the results, showing the position or rank assigned to each photograph by each judge. It will now be possible, by various arithmetical or statistical procedures, to find the amount of agreement among the judges on each of the various traits. Do they disagree more in estimating some traits than in estimating others? On which traits is the agreement greatest? Does the amount of agreement have any necessary relation to the correctness of the judgments? Are some of the judges more representative than others, that is, do their



E



F



G



H

judgments agree more closely with the average or combined rankings by all the judges?

By finding the average position assigned to each photograph it will be possible to learn which faces are judged to indicate the highest degree of the various traits, and which are judged to indicate lower degrees. Is it possible to point out any characteristics that lead judges on the whole to characterize a face as possessing or lacking a given characteristic? If so, what explanation of such tendencies are you inclined to suggest? Do these tendencies have any bearing on the actual significance of such signs?

In the case of Intelligence it is possible to compare the verdicts of the judges, either separately or in combination, with the actual facts. All the individuals whose photographs are here given have been given intelligence tests, and it is possible to assign to each the actual Mental Age and also the Intelligence Quotient (mental age divided by chronological age, in all these cases, since all are less than sixteen years of age.) Compare the individual judgments of Intelligence, and also the combined judgments, with the Mental Ages of the children. What resemblance is found between judgments and this measure of Intelligence? Having computed the Intelligence Quotient of each child, compare the judgments, individual and combined, with these measures. What resemblance is found?

How reliably or accurately is Intelligence, in this instance, estimated from the photographs? Does combination of the verdicts of several judges lead to higher validity than that characterizing the judgments passed by individual judges? How do the results compare with those given in Chapter III?

DATA FOR EXPERIMENT WITH PHOTOGRAPHS

Photograph*	Chronological Age	Mental Age
A	10.0 years	9.0 years
B	12.0 "	9.7 "
C	12.0 "	10.0 "
D	15.0 "	9.8 "
E	12.0 "	8.6 "
F	14.0 "	7.6 "
G	4.0 "	5.7 "
H	5.0 "	7.2 "
I	10.0 "	12.0 "
J	10.0 "	10.8 "
K	7.0 "	6.8 "
L	16.0 "	12.5 "

* These photographs and data are reproduced, with the permission of Prof. Rudolph Pintner, from his article entitled, "Intelligence as Estimated from Photographs," in the *Psychological Review*, Vol. XXV, No. 4, July, 1918, pp. 286-296. After having performed this experiment, the reader should read Pintner's interesting article. The present writer desires hereby to express his thanks to Prof. Pintner and to the *Psychological Review* for their gracious consent to the use of these materials in this book.



I



J



K



L

EXERCISE III

Judging One's Own Characteristics

This experiment requires the frank and conscientious coöperation of all the members of the class or group. There should be at least ten individuals, and fifteen makes a still better number. Of course the larger the group the more reliable the results, but the more laborious the mathematical work involved. Such an experiment honestly conducted with results kept confidential so far as individual cases are concerned, can be made to throw interesting light on individual character.

Select a few interesting trait-terms, such as Intelligence, Aggressiveness, Poise or Balance, Humor, Coöperativeness, Energy, etc. Agree upon a concise definition of each trait-term, and discuss these definitions, with illustrations, so that all the individuals have as nearly as possible the same working conception of what the trait-term is to denote.

Assign to each individual a letter designation, as A, B, and so on. Each individual prepares his own report sheet, but hands it in, marked not with his own letter but with some secret key, known only to himself. This is for indentification and return of the report sheets at a later point.

Each individual then ranks in order of merit, for each of the traits chosen, all members of the group, including himself. Place the whole array of individuals

in a complete order, assigning no two persons to the same rank. Call the best in the trait Number 1, the next best Number 2, etc. The report sheets may now be collected and a table of results prepared, showing for each individual the ranks assigned him by all the members of the group. The median or average rank given to each individual, or the total of all the ranks assigned him, now makes it possible to arrange all the members of the group, for each trait, in a Final Order, based on the combined verdicts of all the judges.

Each individual now recovers his report sheets. In the Final Order he has been given a rank for each trait, and this rank indicates the position in which he stands, according to the combined judgments of his associates. Considering each trait, the individual may now discover how far his own estimate of himself departs from the position assigned him in the Final Order. How accurately does he judge himself in each trait? Is he more nearly correct in some traits than in others? Does he tend to over-estimate or to under-estimate himself? Is this direction of error common to all traits, and fairly equal in amount, or does the constant error of self-estimation vary with the trait? By tabulating such results for all the individuals and taking averages of the tendencies, the general results for all the individuals of the group may be derived. How do these general results compare with those described in Chapter IV?

If Intelligence was included among the traits judged, intelligence tests may be taken by all members of the group and their scores recorded. All individuals may be ranked in order on the basis of these scores. How do the individual's judgments of himself accord with the

results of the objective tests? How accurately do individuals estimate their own intelligence, as compared with other individuals known to them?

Carefully preserve all these records for use in a later experiment.

EXERCISE IV

The Personal Interview

Decide upon some appointment or position, familiar in its general character to all members of the class, for which candidates are to be interviewed. Such positions as the following may be suitable for college classes: editor of the college paper; student adviser to Freshmen entering the college; business manager of the athletic organization; delegate to an inter-collegiate conference on some contemporary question of college life; recipient of a prize to be awarded to the best all-round college student. Current affairs in the college will readily suggest some definite and interesting position of this general character.

Select from the Freshman class ten or twelve individuals who are not personally known to the members of the class, and secure their consent to participate in the experiment. Each member of the class now constitutes himself an interviewer, and interviews each of the applicants in turn, observing the following conditions:

Separate booths or rooms should be available for each interviewer. If four minutes are allowed for each interview, the whole series can be conducted within one hour or one class period. At a signal each interview promptly terminates and the applicant passes to the next interviewer, and so on in turn, until he has visited each interviewer.

Each interviewer is allowed to conduct his interview in whatever way seems best to him, but at the end of each interview he must record his impressions of the qualifications of the candidate in such a fashion that at the end of the hour he can assign to each candidate some score, or grade, or rank, which will make it possible to arrange all the applicants in an order of merit on the basis of his impressions. Care should be taken to avoid giving two candidates the same rank. If it seems desirable to the instructor, some definite system of grading or scoring may be agreed upon, and adopted by all interviewers.

The results from all interviewers may now be compiled in tabular form, and a study made of the agreements and disagreements. A Final Order may be derived, from the combined verdicts, and each interviewer's estimates may be compared with this Final Order. How closely do the interviewers agree? Are some interviewers more representative than others, that is, do their judgments accord more nearly with the Final Order?

Each interviewer, or particularly representative or non-representative interviewers, may now be called upon for an account of their interview method. Upon what criteria did they base their judgments? What influences determined their verdicts? What special difficulties were encountered in attempting to make such ratings of other individuals on the basis of brief interviews.

What indications are there that combining the verdicts of several interviewers gives more reliable results than those secured from single interviewers chosen at

random? If some judges are more expert or representative than others, is there any way of finding out how they come to possess this expertness? What are the indications in the present instance?

Preserve all records for use in a later experiment.

EXERCISE V

Recommendations and Testimonials

In Exercise 3, which has already been performed and recorded, each individual estimated not only his own traits, but also judged the degree to which other persons of his acquaintance possessed these same traits. The data already accumulated in Exercise 3 may now be made the material for a study of the judgments of associates.

Each member of the class received a Final Rank, based on the combined judgments of all the members of the class. Taking each individual in turn, for a given trait, determine the Mean Deviation of the judges from this Final Rank. Do the judges differ more in estimating some individuals than in estimating others? Why? Determine the degree of deviation in this manner for all the traits. Averaging the Mean Deviations gives a measure of agreement of judges on each trait. Do the judges agree more closely in estimating some traits than in estimating others? How do these results compare with those given in Chapter 6? If a hierarchy of agreement is shown, how does this hierarchy compare with those suggested in Chapter 6?

Each individual should now classify all members of the class, except himself, on the basis of the degree of his acquaintance with them. Three degrees may conveniently be used, such as Intimately Acquainted, Fairly

Well Acquainted, Only Slightly Acquainted. Consider now your own estimates of these individuals, and their degree of deviation from the Final Order assigned them. Does degree of acquaintance affect the accuracy of your judgments of your associates?

Each individual's judgments of the group may be correlated with the Final Order, thus giving a measure of his judicial capacity, or his representative character, or his agreement with the consensus of opinion. How do individuals differ in their "judicial capacity" as thus described? Does the individual whose verdicts are representative in one trait also tend to show corresponding judicial capacity in other traits?

Individuals may now be ranked according to their "judicial capacity" in a given trait, and also according to their position in the Final Order for this trait. Is there any relation between possession of the trait in high degree and ability to judge that trait in others? How does this vary with the trait, and upon what factors do the results apparently depend?

Compare the abilities of all the individuals in (a) estimating their own traits; (b) estimating the traits of strangers in the interview; (c) estimating the traits of associates. Are the representative judges in the one case also representative in the other cases? What correlations are found? Is there such a thing as general judicial capacity in this connection?

Is there any relation between the ability to judge according to the consensus of opinion and the scores in the intelligence tests?

What further questions of interest might be raised and answered, for the circumstances in question, on the basis of the data now at hand?

EXERCISE VI

Rating Scales

The following experiment illustrates the use of the "concrete specimen" rating scales, without introducing the factor of personalities. The same general principle may be utilized in judging people, and, at the discretion of the instructor, the method may be adapted to rating persons. Modifications of the method which involve the principles of the "graphic" rating scales may also be introduced if time allows.

Rating Works of Fiction.—Each individual suggests the qualities which a work of fiction, as a novel, should possess. Through discussion, several essential qualities are agreed upon. Several novels with which all the members of the class are acquainted are then chosen for rating.

Each individual rates each book, for each of the qualities, by what seems to him to be the best method—thus he may assign letter grades, per cent grades, descriptive terms, etc. The various ratings for a given trait are then compared, and the difficulties and ambiguities encountered by the use of such heterogeneous methods are exhibited. Attempts may be made to summate the grades assigned to a given book, in its various qualities, and the difficulties of such summation on the basis of these methods of grading will be revealed.

Each individual now prepares, for each trait, a con-

crete specimen scale, after the following manner: Suppose the quality is Realism, this being defined, by agreement, to mean "fidelity of detail to the real conditions of the epoch in which its scenes are laid." Select the most realistic novel of your acquaintance, and call it A. Now select the least realistic novel of your acquaintance and call it E. Select some novel, C, which stands just half-way between A and E in this quality. Now select B which stands mid-way between A and C, and select D which stands mid-way between C and E Write down the names of these novels, under the heading of the trait in question. Make similar scales for each of the qualities to be considered, writing out the names of the books under the appropriate quality headings. The same books need not of course appear in all the scales. Each scale is to be made up independently of the others, using for each quality all the books you know, as possible choices.

Now turn to the books to be graded by all the members of the class. Taking each book in turn, locate it in the appropriate scale for each quality, as for example, "as good as C," "half-way between D and E," etc. It may be well to assign a value of 5, 4, 3, 2, 1 to A, B, C, D, E, respectively.

Consider now the gradings of a given book made by all the individual judges. Is agreement, consistency, or clearness facilitated by the use of these concrete specimen scales? Secure the total value of each book, for your own rating, by summating the numerical values for all the qualities. How do these totals differ from individual to individual and from book to book? Distributing such scores in a surface of frequency, what distribution is observed?

If the various qualities are not equally important, how might you proceed to assign appropriate weights to the various qualities, so as to secure more accurate total scores or summation values for each book?

What special difficulties or possible sources of error arise in connection with the use of such rating scales? How eliminate them?

EXERCISE VII

General Principles of Tests

The instructor may select, from the general equipment of the laboratory, a typical array of tests. Each test is administered to the class, and the method of scoring and evaluating the results is discussed. In this connection special attention may well be given, not so much to the results of the tests, but rather to the general principles of construction, expression, and standardization employed.

Faulty tests should be exhibited, which reveal defects in materials, in instructions, in technic, in scoring, in interpretation, etc. In this way the difference between a "test" and an "instrument of measurement" may be made clear.

In the case of each satisfactory test, consider first the method of construction—does it exemplify the Standard Task, Homogeneous Material, Graded Tasks, Miscellaneous Gradations, Response Values, or some combination of these, or some new principle, or some ambiguous principle?

Consider now the method in which the results of the test are expressed. Is the method that of Original Scores, Percentile Units, Distribution Units, Developmental Units, or Absolute Units? Are any new or ambiguous principles revealed?

Students may be assigned the task of planning the development of a new test on one or more of the principles of construction. In such cases the steps necessary before the test assumes its final form may be outlined or described.

EXERCISE VIII

Measuring General Competence

A study of various tests or test series for the determination of "general intelligence" may be made at this point. For general class purposes it is more convenient to use tests that are suitable for group procedure. It should ordinarily be made clear that such a demonstration as that here described does not constitute adequate training in the use of such tests, but represents merely a demonstration of some of the more commonly used forms, and a comparison of them.

After several forms of intelligence tests have been used with all the members of the class, the scores in each test may be computed, and comparisons made between the various sets of results. In this connection the possible results or influences of familiarity, practice, differences of age, education, previous occupation, etc., may be considered.

The Army Alpha Test will be one suitable form for adults. The scores achieved by members of the class may be distributed in a surface of frequency, and a study made of their range. In "Army Mental Tests," by Yoakum and Yerkes (Henry Holt & Co.), may be found various tables of typical results, showing scores for various occupations, colleges, college classes, ages, races, and similar data. This book also contains instruc-

tions for giving and scoring the Alpha tests, and samples of the forms to be used.

The Otis Group Tests may also be used to advantage. Although they are in general not difficult enough to afford actual differentiation of college students, this fact in itself emphasizes certain important facts about intelligence tests and their use, the advantages and limitations of various forms, and other related points.

Several of the tests described in "The Mental Survey" by Pintner may be used to advantage, as special tests, each of which is separately standardized.

In "The Psychological Study of College Students" by Carothers (*Archives of Psychology, Columbia University*) are given standard instructions and norms on college freshmen in Barnard College. Some of these tests may be adapted to group procedure, but in general it will be better to leave these tests for a later experiment on psychographic methods.

If the instructor should desire to illustrate some method of individual examination, the Stanford Revision of the Binet-Simon Tests may be given by him to one or more members of the class, after the manner of a demonstration. The scores of the individuals examined by this method may be compared with the scores received by the same individuals in the various group tests.

The various advantages and disadvantages of group tests as contrasted with individual examinations may be considered in detail. Special methods for the mental examination of the illiterate, the foreigner, the blind, and the deaf may also be considered and illustrated.

EXERCISE IX

Measurement of Special Aptitudes

The measurement of special aptitudes, as distinguished from general competence or intelligence, may conveniently be illustrated by employing the following materials and methods.

a. Measures of Musical Talent.—The Seashore tests for musical abilities, including Sense of Pitch, Sense of Intensity, Sense of Time, Sense of Consonance, and Tonal Memory. These may be secured in the form of phonograph records from the Columbia Graphophone Company. A brief manual of instructions is also provided, which contains norms for adults and for 8th grade and 5th grade children.

b. Mechanical Ability.—The Stenquist Assembling Tests, which are furnished by C. H. Stoelting Company, Chicago, afford interesting measures of the ability to handle simple mechanical constructions. Manual of directions is issued along with the tests, and includes methods of scoring, norms, and other necessary information.

c. Kelley Construction Test.—This is furnished by Stoelting, along with directions. A set of stereoscopic slides, containing the scale for grading is also required.

d. Motor Tests.—Various tests of motor capacity, such as Steadiness, Rate of Tapping, Coördination, and so on, are described in *Manual of Mental and Physical*

Tests by Whipple (Warwick and York). Useful tests to employ for this purpose are Grip, Tapping and Coördination. Instruments required for these tests are furnished by Stoelting.

e. Sensory Tests.—Standard tests for Visual Acuity, Auditory Acuity, Color Discrimination, and similar abilities or processes are described in Whipple's Manual. Tests of Visual Acuity, and of Color Discrimination may be readily used if the necessary materials or instruments are available.

To give all these tests to all the members of a class would require a very great deal of time. Selections should be made from the list or from other suitable materials, and typical illustrations given. It will be found instructive to compare, by correlation or otherwise, the scores attained in these special tests, with those made in the tests of general intelligence. The special tests may also be correlated with one another. In this manner the special character of these tests will be emphasized.

EXERCISE X

Measures of Knowledge and Skill

The Trade Tests may be used to advantage in illustrating the method of the Sample. Typical trade tests are given in "Trade Tests," by J. C. Chapman (Henry Holt & Co.). These were developed by the Trade Test Division of the U. S. Army. Some of them relate to occupations which may be more or less familiar to some members of the class, and these occupations may be selected for the purpose of this exercise. In general the following tests will show some fairly high scores among most college classes. In giving the tests, follow closely the printed instructions for administration and scoring:

Oral Trade Tests	Picture Trade Tests	Performance Trade Tests
Tailor	General Carpenter	These are not suited to general laboratory use, but may be described. The test for Typist may be tried out.
Butcher	Cobbler	
General Carpenter	Harness Maker	
Horseshoer	Horseshoer	
Auto Mechanic		

In considering the scores in these various trade tests, is there any correlation between the various tests? How are the facts to be explained? Are there correlations between scores in these trade tests and scores previously made in intelligence tests? How explain the facts?

Make a careful study, in Chapman's book, *Trade Tests,* of the procedures followed in developing and standardizing these trade tests. Now choose some occupation with which you are familiar and which finds many representatives in the territory in which you are living. Describe the way in which you would go about developing a Picture Trade Test for this occupation. Outline the various steps in detail, indicating the probable difficulties you would encounter, and the manner in which these difficulties should be met.

Following the method described for the development of Oral Trade Tests, undertake the collection of a set of questions for the occupation you have chosen to study. Carry out the steps of preliminary and final testing of the questions, calibrate the test, and assemble it in final form. Such occupations or trades as the following will provide interesting opportunity for such an exercise: Cook, chamber-maid, elevator boy, janitor, laundress, nurse, teamster, barber, policeman, reporter.

From your study of the occupation you have chosen, what would seem to be the relative advantages and disadvantages of the three chief forms of trade tests—oral, picture and performance? Why?

EXERCISE XI

Educational Measurement

1. *Drawing.*—Each student makes as good a drawing as possible of either a Church, a Person or a Snow-fight. Each then examines the drawings of all and assigns to each specimen, without consultation with others, a rating or grade in terms of general excellence. Compare the ratings given by different judges to the same picture. How do they agree, how intelligible, precise, and quantitatively comparable are they, how exact and definite as measures of drawing excellence?

Each student is now provided with a copy of Thorndike's Drawing Scale, and rates each specimen on this scale, following the standard instructions. Ratings on this basis are now compared. What are the advantages of such a Product Scale? How might this one be improved? To what practical uses might such a scale be put?

2. *Handwriting.*—Specimens of handwriting are provided, and these are graded first on the ordinary per cent method commonly used in grading school exercises. Some standard handwriting scale is then used, and the specimens measured by the use of this scale. The advantages and practical uses of such a scale may then be considered, in the light of the experimental results.

3. *Reading.*—Copies of the Burgess Reading Scale or of the Thorndike-McCall Reading Scale are provided,

and all the members of the class are put through the test for the measurement of ability in silent reading. What particular problems come up in connection with the use of such scales and the interpretation of results? A comparison of scores in these tests may be made with the scores in the intelligence tests previously accumulated. What relations are suggested?

4. *Sewing.*—The Murdoch Sewing Scale may be exhibited and its method of construction studied.

5. *Composition.*—The Hillegas Composition Scale may also be studied to advantage, especially from the point of view of the principles involved in its construction. If time allows, each student may be asked to prepare a composition suitable for measurement on such a scale. These compositions may then be graded in the usual fashion and these grades compared with ratings made when the scale is used.

6. Considering the educational scales that have now been studied, consider them from the point of view of the principles of construction and the principles of expression described in Chapter VIII. Which principles are illustrated in these educational scales? Are any new principles revealed?

EXERCISE XII

The Correlation Methods

In the first Table are given data concerning eighteen salesgirls in a department store. For each are given the following: Age, Years of Selling Experience, Weekly Salary, Average Weekly Bonus Earned, Education, Manager's Estimate of Sales Ability (Excellent, Good, Fair, Poor), Salesmanship Instructor's Estimate of Sales Ability (Rank in the Group), Instructor's Estimate of Intelligence (Rank in the Group). In some cases several individuals are given the same rating, indicating estimated equality. Sign + means somewhat better than the term assigned.

While these data leave much to be desired for the purpose of determining an objective criterion of sales ability, they represent all the information available in the actual case. Suppose that it is desired to find tests that indicate sales ability, for the selection of new applicants. The first step will be to find tests that correlate well with such ability in the case of employees actually on the job. In order to correlate or otherwise compare test scores and sales ability, some more or less objective criterion of sales ability must first be secured. Use any of the data given, or all. Treat each item separately or combine some or all of them. It will be well to discuss the general problem first, and then for different members of the class to adopt different sorts of objective criteria.

DATA CONCERNING SALES GIRLS

Name	Age	Years of Selling	Weekly Salary	Average Weekly Bonus	Manager's Estimate	Instructor's For Sales Ability	Estimate For Intelligence	Education
A	32	15	$9.00		Fair	7	6	8th Grade
B	26	8	10.00	$2.00	Excellent	3	15.5	7th "
C	17	0.3	8.00		Excellent	3	15.5	7th "
D	20	2	8.50	.75	Good	12	10	8th "
E	21	0.5	8.00	.75	Poor+	12	8	8th "
F	29	11	13.00	3.00	Good	11.5	15.5	7th "
G	19	1	7.00	1.00	Good+	3	12	8th "
H	24	2	10.00	.75	Good	12	11	8th "
I	18	0.5	8.00		Fair	7	4	7th "
J	18	2	9.00	.50	Good	12	9	7th "
K	21	4	9.00		Good	17	15.5	7th "
L	21	4.5	9.00		Good	18	18	6th "
M	23	5.5	10.00	3.00	Excellent	7	7	8th "
N	23	5	10.00	.50	Fair	12	4	8th "
O	49	5	10.00		Good	3	1.5	2nd H. School
P	25	0.3	12.00		Excellent	3	1.5	3rd College
Q	18	2	7.00		Excellent	16	13	7th Grade
R	32	10	9.00		Fair	12	15.5	8th "

The results in the different cases may then be compared after the next step is completed.

In Opposites, Substitution, Analogies, Word Reconstruction, and Cube Imitation the score is the amount of work or number of items accomplished under fixed conditions of time or trials. In Color Naming the score is the time required for a fixed task.

Scores of Salesgirls in Several Tests

Name	Naming Opposites	Giving Analogies	Word Reconstruction	Substitution	Color Naming	Cube Imitation
A	30	9	2	89	66.2	4
B	33	15	4	86	67.8	4
C	31	19	4	95	55.2	7
D	24	20	5	68	60.8	5
E	21	18	4	76	84.2	6
F	34	8	3	33	66.0	4
G	26	13	5	70	80.0	4
H	23	8	6	50	71.2	6
I	40	36	6	83	56.0	9
J	38	22	8	100	57.0	7
K	20	14	3	89	55.5	6
L	15	7	3	75	81.0	6
M	21	13	3	72	81.5	5
N	29	15	5	77	75.5	7
O	26	31	11	84	55.4	7
P	40	32	6	55	68.0	8
Q	12	6	1	77	71.8	5
R	21	10	3	60	57.5	6

In the preceding table are given the results when several tests were applied to these individuals. In each case the individuals are given their scores in each test.

By comparing these tests results with the chosen criterion derived from the foregoing table, select a team of tests that seem to indicate best of all the relative sales ability of these individuals. With the advice of the instructor, various methods of comparison may be employed.

Do you find that some of these tests are better indicators of sales ability than are others? How would you select the best team of three tests? Would you give each of these three equal value in computing the total score of an individual? If not, how would you weigh the various tests?

Look up, in books on statistical methods, the methods for partial and multiple correlation, and indicate the way in which these methods might be put to significant use in such an investigation as this.

What seems to you, from your experience with these results, to be some of the principal difficulties encountered in the application of the correlation methods to problems of employment and placement?[1]

[1] The data used in this exercise were secured in a study conducted by Elsie Oschrin Bregman. After completing the exercise it will be interesting to read "Vocational Tests for Retail Saleswomen," by Elsie Oschrin, in *Journal of Applied Psychology*, June, 1918. For further examples consult *Employment Psychology*, by H. C. Link, and various other references given in the Bibliography preceding this Appendix.

EXERCISE XIII

Temperamental Characteristics

The following experiments, selected from the many available ones, may readily be planned for group or individual experiments. The results in each case may be compared with the results of preceding tests and exercises, as with self-estimates, judgments of associates, intelligence tests, tests of special aptitudes, and so on. In general the plan will be to become familiar with the proposed method by actual use of it, and to discuss its merits and shortcomings.

1. *Downey Will Profile.*—For the determination and measurement of the volitional pattern. Necessary blanks, instructions, materials and norms, either for the individual form or for the group from of this test, are to be secured from the World Book Company, Yonkers, New York.

2. *Moore Aggressiveness Test.*—A description of the methods and results of this series of tests is to be found in "The Measurement of Aggressiveness," by H. C. Moore, *Journal of Applied Psychology,* June, 1921.

3. *Kent-Rosanoff Association Test.*—Instructions and lists of stimulus words, with table of results are to be found in "A Study of Association in Insanity," by Kent and Rosanoff, *American Journal of Insanity,* Numbers 1 and 2, 1910. The method is also described in *Whipple's Manual of Mental and Physical Tests,* Vol.

2, Test 33A, page 53, with a full discussion of the results secured by this method under various circumstances.

4. *Suggestibility Tests.*—In Chapter X of Whipple's Manual of Mental and Physical Tests five tests are described and discussed in detail. The materials needed for these tests are furnished by C. H. Stoelting Company, 3037 Carroll Avenue, Chicago. See also "Individual and Sex Differences in Suggestibility," by Warner Brown, *University of California Publications in Psychology,* Vol. II, No. 6, July, 1916.

5. *Tests of Mental Balance.*—Consult Chapter VII of Wells, *Mental Adjustments* (Appleton, 1917), for accounts of various methods of attempting to measure mental balance and moral perception. Fernald's test, or some adaptation of the original experiments reported by Wells, may be used as a class exercise.

EXERCISE XIV

Psychographic Methods

1. In the foregoing exercises many measurements of each member of the class have been secured. Some of these are not directly comparable with the others, but several of them can be brought together for purposes of psychographic presentation and analysis. Thus the Intelligence tests, the tests of Musical Ability, of Mechanical Ability, the Motor and Sensory tests, the Will Profile, the Association Test and some of the Suggestibility tests, all afford records which can be expressed in approximately comparable form. Select some form of expression, as Mental Age, Percentile, or Distribution Units, and express your own records in each test in this form.

Now construct a psychographic chart similar to those illustrated in Chapter XII, and record thereon your own standing in each of the tests you have selected. This will give a partial profile or psychographic record of your abilities and characteristics. What in general does the chart tell you about your mental equipment as shown by the tests? Compare your own profile with those of other members of the class. Discuss the differences and resemblances found. To what practical use might such data be put? To what extent do they represent a complete analysis of your personality?

2. Some of the general test systems, as the Alpha test

for example, consist of several different tests which may be independently scored and which seem to involve somewhat different types of work. The total score in such a test system may be analyzed into its component scores, and these partial scores may be represented in psychographic form. The article by Yerkes and Cobb, given in the Bibliography for Chapter XII, may be consulted in this connection as an interesting illustration of the method.

3. In Carothers, *Psychological Study of College Students,* will be found an array of tests described, from the Woodworth-Wells Series. Instructions for giving, scoring, and evaluating each test are given, and norms on college students are also presented. Each member of the class may be given this array of tests and a psychographic record and analysis made of the results. Test blanks for this series may be secured from Stoelting.

EXERCISE XV

Examination of an Individual

A stranger, either child or adult, may be brought before the class for examination, and a very elaborate experiment planned. This experiment may of course be abbreviated according to the time available.

1. On the basis of the appearance and manner of the subject each student records his impression of the individual's intelligence, and of such other traits as may be agreed upon. These records may later be compared with one another and with the results of the examination.

2. Each student may be allowed to interview the subject, briefly, and record his estimates of the individual's characteristics on the basis of this interview.

3. Selected intelligence tests may be given the subject, in the presence of the class, preferably an array of tests each of which can be independently scored and recorded. One of the miscellaneous scales as the Stanford-Revision, may also be used for a summary intelligence score.

4. Tests of special aptitudes, such as the Musical Talent tests, the Stenquist Construction Test, and some of the Motor and Sensory tests should also be employed.

5. Educational Tests, for which there are norms which make the results comparable with the other measure-

ments, may also be used if time permits or if the subject can appear on several occasions.

6. Selected temperamental tests, as for example the Will Profile, The Kent-Rosanoff and some of the Suggestibility tests may also be given with profit.

On the basis of these results, each student, or the class in joint enterprise, now prepares a psychographic record of the results of the complete examination. Discussion may center about the significance of such an intensive study of the individual, and the practical uses to which the results may be put.

This experiment requires, in the full form here described, a great deal of time, care, patience, and expertness. If only one afternoon or forenoon is available, the experiment should be abbreviated in such a fashion that whatever measurements are made may be made under favorable conditions and may conform to the most rigorous standards of technic.

EXERCISE XVI

Installation of Employment Supervisor

Assume that you either own or have direct charge of a business which has grown to such proportions that it seems desirable to establish, at least in a preliminary way, a special department of personnel direction. In your capacity as owner or director you have just employed a college graduate of high general intelligence, a general familiarity with your line of business, and a personal interest in people. He is to take charge of the employment and placement of all the factory operatives or of all the clerical personnel.

1. What directions would you give him in connection with the use of letters of application and the development of application forms, so that these materials should afford him the greatest amount of reliable data, for his guidance?

2. What practice or training would you prescribe in order to ensure the correspondence of his judgments with what experience has shown you to be the best general policy in the selection of individuals and their placement in the organization?

3. What devices and methods would you recommend to him, in order to help him get the maximum amount of reliable information from letters of reference and testimonials?

4. What general principles, useful suggestions, and particular cautions would you give him in order to improve his interviewing of foremen and supervisors regarding job requirements?

5. What helpful suggestions would you give him in order to increase the accuracy and effectiveness of his personal interviews with candidates for employment?

6. What special sources of error should he learn to take into account in evaluating the testimony of others or in rendering judgments of his own, concerning the characteristics of other people?

7. If the workers whom he is to select and place should have fairly definite degrees of general intelligence, what group tests might he well undertake to have in readiness for the examination of applicants in this respect?

8. How in general should he proceed to select special tests which would make it possible to determine whether a given applicant should, for most efficient and satisfying work, be assigned to the group of clerical workers or to the group of factory operatives?

9. If it should seem best for him to pursue a special course of instruction or training in vocational and industrial psychology or in personnel methods and human engineering, through home study, prescribe for him a series of books dealing in a valuable way with such topics as—statistical methods, employment practice, vocational psychology, character study, mental measurement, social psychology, industrial management, job analysis, labor problems, wage adjustments, office management, business organization. If you are not familiar with these fields consult a reference librarian or specialists in the various fields.

INDEX

Absolute units, 118
Aggressiveness, 188
Allen, 176
Allport, 57, 216
Analogies, 124
Analogy, method, of 152
Analysis,
— of individuals, 205
— of interviews, 71
— of jobs, 201
— of self, 45, 231
— vocational, 142
Anderson, 43
Andrews, 29
Appendices, 215, 225
Application, letters of, 9, 226
— analysis of, 29
Applications, judgments of, 10 ff.
— improvements in, 24
Aptitudes, 139 ff., 249
Association, method of, 191, 196
— tests of, 128
Attitudes, 139 ff.
Aviators, tests of, 30, 176

Bagby, 176
Balance, 183, 191, 204
Bibliography, 215 ff.
Bills, 175
Binet, 66, 125, 186
Binet-Simon tests, 126
Blacksmith, tests for, 156
Brown, 196, 257
Burtt, 176

Carney, 176
Carothers, 245, 259
Cattell, 77
Chapman, 175, 248
Character, analysis of, 161, 181, 228, 259
— judgment of, 2
— meaning of, 2
— types of, 191
Clerical workers, tests of, 162
Cobb, 211
Competence, general and special, 139
— measurement of, 112, 244
Complexes, 192
Comptometers, tests of, 175
Consistency, hierarchy of, 78
— promotion of, 102
Construction, of tests, 116
Correlation, meaning of, 16
— method of, 152, 161, 252
— of judgments, 17, 22, 38, 40, 56, 95
— of tests, 143, 166, 172, 252
Courtis, 211

Demonstrations, 226 ff.
Developmental units, 120
Distribution units, 118

Dockeray, 176
Domestic worker, form for, 108
Downey, 186, 210, 256
Duck, 175

Educational measurement, 158, 253
Emotion, 196
Employees, tests of, 143, 166, 171, 172
Engineers, tests of, 176
Errors, of judgment, 48 ff.
— sources of, 91
Estimates, of self, 45 ff.
Exercises, 226 ff.
Experiments, with employees, 166
— with interviews, 62
— with letters, 12, 20
— with photographs, 34, 42
— with self-analysis, 48
Expression, methods of, 117
Eye-movements, 189

Folsom, 83
Fontegne, 176
Form, a model, 108
Fortune telling, 5
Franzen, 54

Gates, 131
Graded tasks, 116
Graphology, 185
Group tests, 135
Guidance, 141

Handwriting, measurement of, 129
— significance of, 185
Head, 176
Healy, 147
Henmon, 176
Hierarchy, of consistency, 78
Hillegas, 251
Hollingworth, 5, 175
Homogeneous materials, 116
Hoover, 176
Hull, 185

Individual tests, 135
Inquiry forms, 108
Inspectors, tests of, 171
Instability, 204
Intelligence, correlations with, 35, 57, 98, 174
— in photographs, 35, 42
— judgments of, 35, 57, 98
— tests of, 140
Interview, methods of, 72
— reliability of, 62, 237
Inventory, psychoneurotic, 183
Irle, 94
Isaacs, 176

Job analysis, 201
Jones, 176
Judges, qualification of, 55
Judgment, and perception, 1
— consistency of, 16
— data of, 4
— errors of, 91
— methods of, 8
— of self, 45
— reliability of, 20, 36
— variability of, 15, 34, 49, 74, 79, 85
Jung, 191

Kelley, 246
Kent, 128, 261
Kitson, 211
Knight, 54
Kohs, 94
Kraepelin, 212
Krais, 176

Laboratory exercises, 225 ff.
Lahy, 175
Letters, of application, 9, 226
Link, 73, 171, 175, 181, 255
Lipmann, 176
Lough, 175

Magic, 3
Mann, 87
Marcus, 175
McCall, 250
McComas, 176
Measurement, in education, 158, 253
— of aggressiveness, 188
— of handwriting, 129
—of suggestibility, 195
— of will pattern, 186
Mechanical ability, 148
Methods, of construction, 116
— of expression, 117
— of measurement, 112
Miner, 81, 182
Miniature, method of, 151
Miscellaneous gradations, 116
Montgomery, 185
Moods, 194, 196
Moore, 188, 256
Moral traits, 181
Motor men, tests of, 176
Muensterberg, 176
Murdoch, 251
Music, 148

Nead, 176
Norsworthy, 77

Operators, tests of, 176
Organism, quality of, 96
Original units, 117
Oschrin-Bregman, 172, 175, 255
Otis, 147, 176, 245

Percentile units, 117
Photographs, character in, 33
— judgments of, 34, 228
— value of, 39
Picture tests, 157
Pintner, 42, 147, 230, 245
Poffenberger, 20, 175
Pressey, 197
Printers, tests of, 176
Production records, 165
Profile, 186, 201, 205, 210
Psychiatry, methods of, 100
Psychograph, 201 ff., 258
Psychoneurotic inventory, 183

Qualities, temperamental, 145

Rating scales, 103, 182, 239
Recommendations, value of, 74, 237
Records, meaning of, 130
— of production, 162
Response values, 116

Rice, 149
Rogers, 166, 175
Rosanoff, 128, 256
Rossolimo, 210
Rugg, 104, 110

Sachs, 176
Sales clerks, tests of, 172, 176
Salesmen, study of, 26
Sample, method of, 151
Scales, for rating, 103, 239
Scattering, 204
Scott, 68, 104, 176, 182
Seashore, 148, 176, 211
Self-analysis, 45, 48, 231
Simon, 125
Slawson, 87
Solari, 176
Sommer, 213
Stamm, 176
Standards, of report, 99
Standard Task, 116
Stenographers, tests of, 166
Stenquist, 148, 246
Stern, 176, 195, 210
Stratton, 176
Suggestibility, 195

Technic, improvements in, 23, 58, 69, 109
Temperament, 145, 256
— diagnosis of, 179 ff.
Tendencies, of judgment, 91
Terman, 218
Testimonials, value of, 74, 97, 237
Tests, and measures, 113
— choice of, 131, 162
— construction of, 116
— correlations of, 143
— meaning of, 130
— nature of, 114, 242
— of employees, 142, 154, 164, 166, 171, 172
— of interviewers, 62
— of judgment, 11, 20
— of recommendations, 74
— of temperament, 179 ff.
— samples of, 119 ff.
— types of, 116, 135, 147, 242
Thorndike, 30, 93, 147, 174, 176, 250
Thurstone, 175
Toulouse, 201
Trabue, 147, 176
Trade tests, 148, 154, 248
Traits, objective and subjective, 78, 84
Types, of association reaction, 192
— of character, 191
Typists, tests of, 166

Units, of measurement, 117

Variability, elimination of, 100
Vartanian, 20
Vocational analysis, 151

Washburn, 196
Wells, 26, 147, 193, 250
Whipple, 195, 247
Will profile, 186
Woodworth, 147, 183

Yerkes, 211
Yoakum, 244

RETURN TO ⟶ **EDUCATION-PSYCHOLOGY LIBRARY**
2600 Tolman Hall 642-4209

LOAN PERIOD 1	2	3
1 MONTH	QUARTER LOAN	
4	NO TELEPHONE RENEWALS	6

ALL BOOKS MAY BE RECALLED AFTER 7 DAYS
2-hour books must be renewed in person
Return to desk from which borrowed

DUE AS STAMPED BELOW

APR 0 6 REC'D 12 PM

UNIVERSITY OF CALIFORNIA, BERKELEY
FORM NO. DD10, 5m, 3/80 BERKELEY, CA 94720

U.C. BERKELEY LIBRARIES
C029517358